GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 12, 2004, 15:21:09 ; Search time 41.6471 Seconds
(without alignments)
400.276 Million cell updates/sec

Title: US-09-620-955B-10
Perfect score: 287
Sequence: 1 LVPRGSMATLEKLMKAFESL.....QQQQQQQQQLQPGSTRAAAS 59

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	287	100.0	59	4	AAB69605	Aab69605 Huntingti
2	231	80.5	66	4	AAB69613	Aab69613 Huntingti
3	218	76.0	145	4	AAB69614	Aab69614 Huntingti
4	208	72.5	63	5	AAE26651	Aae26651 Human hun
5	208	72.5	64	4	AAB69607	Aab69607 Huntingti
6	208	72.5	86	2	AAW95073	Aaw95073 GST-HD fu
7	208	72.5	86	2	AAW95078	Aaw95078 GST-HD fu
8	208	72.5	89	4	AAB69608	Aab69608 Huntingti
9	208	72.5	94	2	AAW95075	Aaw95075 GST-HD fu

10	208	72.5	94	2	AAW95080	Aaw95080 GST-HD fu
11	208	72.5	98	4	AAB69610	Aab69610 Huntingti
12	208	72.5	108	2	AAW95071	Aaw95071 Amino aci
13	208	72.5	108	2	AAW95076	Aaw95076 Amino aci
14	208	72.5	121	4	AAB69609	Aab69609 Huntingti
15	208	72.5	123	4	AAB69611	Aab69611 Huntingti
16	208	72.5	155	4	AAB69612	Aab69612 Huntingti
17	208	72.5	171	5	AAE26650	Aae26650 Human hun
18	197	68.6	3223	4	ABB11407	Abb11407 Human Hun
19	197	68.6	3223	4	ABB11470	Abb11470 Human Hun
20	196	68.3	79	4	AAB69616	Aab69616 Huntingto
21	196	68.3	171	2	AAW99022	Aaw99022 Human hun
22	196	68.3	513	2	AAY33500	Aay33500 Human hun
23	196	68.3	530	2	AAY33501	Aay33501 Human apo
24	196	68.3	552	2	AAY33502	Aay33502 Human apo
25	196	68.3	589	2	AAY33503	Aay33503 Human apo
26	196	68.3	3144	2	AAR58777	Aar58777 Protein e
27	196	68.3	3144	2	AAW36887	Aaw36887 Previousl
28	196	68.3	3144	2	AAW09871	Aaw09871 Human hun
29	196	68.3	3144	2	AAW44742	Aaw44742 Human hun
30	196	68.3	3144	2	AAY33493	Aay33493 Human hun
31	181	63.1	87	5	ABG30880	Abg30880 Human pro
32	181	63.1	87	6	ABP72612	Abp72612 Huntingto
33	181	63.1	1542	5	ABB78013	Abb78013 Amino aci
34	180	62.7	55	2	AAW95072	Aaw95072 GST-HD fu
35	180	62.7	55	2	AAW95077	Aaw95077 GST-HD fu
36	180	62.7	63	2	AAW95074	Aaw95074 GST-HD fu
37	180	62.7	63	2	AAW95079	Aaw95079 GST-HD fu
38	172.5	60.1	3139	2	AAY08898	Aay08898 Human Hun
39	152	53.0	69	4	AAB69604	Aab69604 Huntingti
40	147	51.2	1081	6	ABR53539	Abr53539 Protein s
41	147	51.2	1109	7	ADC59312	Adc59312 Human pol
42	147	51.2	1340	6	AAE37017	Aae37017 Human nuc
43	147	51.2	1761	4	ABB59512	Abb59512 Drosophil
44	145	50.5	80	4	AAB69622	Aab69622 TATA bind
45	145	50.5	338	5	AAU77921	Aau77921 Human Tat

ALIGNMENTS

RESULT 1
AAB69605
ID AAB69605 standard; protein; 59 AA.
XX
AC AAB69605;
XX
DT 30-APR-2001 (first entry)
XX
DE Huntingtin accumulation inhibitor peptide GST-HD-Q25.
XX
KW Neurological disorder; Huntington's disease; Alzheimer's disease;
KW Parkinson's disease; prion disease; frontotemporal dementia;
KW amyotrophic lateral sclerosis; spinal and bulbar muscular atrophy;
KW dentatorubal-pallidoluysian atrophy; spinocerebellar ataxia type 1; SCA2;
KW SCA3; SCA4; SCA5; SCA6; SCA7; protein accumulation; intrabody.
XX

OS Synthetic.
XX
PN WO200106989-A2.
XX
PD 01-FEB-2001.
XX
PF 24-JUL-2000; 2000WO-US020131.
XX
PR 27-JUL-1999; 99US-0146047P.
PR 21-JUL-2000; 2000US-00620955.
XX
PA (HUST/) HUSTON J S.
PA (MESS/) MESSER A.
PA (LECE/) LECERF J.
XX
PI Huston JS, Messer A, Lecerf J;
XX
DR WPI; 2001-182700/18.
XX
PT Inhibiting intracellular polypeptide accumulation, useful for treating
PT neurological disorders, e.g. Alzheimer's disease, comprises contacting
PT the polypeptide with a specific intrabody.
XX
PS Disclosure; Page 96; 108pp; English.
XX
CC The present invention describes a method for inhibiting the formation of
CC aggregates of certain proteins, involving contacting the protein with a
CC binding molecule known as an intrabody. Proteins to be bound include
CC those associated with neurological disorders, and so the method can be
CC used in the prevention of diseases such as Alzheimer's, Parkinson's and
CC Huntington's diseases, prion disease, frontotemporal dementia,
CC amyotrophic lateral sclerosis, spinal and bulbar muscular atrophy,
CC dentatorubal-pallidoluysian atrophy, spinocerebellar ataxia type 1
CC (SCA1), SCA2, SCA3, SCA4, SCA5, SCA6 and SCA7
XX
SQ Sequence 59 AA;

Query Match 100.0%; Score 287; DB 4; Length 59;
Best Local Similarity 100.0%; Pred. No. 4.6e-26;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 LVPRGSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQLQPGSTRAAAS 59
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 LVPRGSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQLQPGSTRAAAS 59
```

RESULT 2
AAB69613
ID AAB69613 standard; protein; 66 AA.
XX
AC AAB69613;
XX
DT 30-APR-2001 (first entry)
XX
DE Huntingtin accumulation inhibitor peptide GFP-HD-Q25.
XX
KW Neurological disorder; Huntington's disease; Alzheimer's disease;

KW Parkinson's disease; prion disease; frontotemporal dementia;
KW amyotrophic lateral sclerosis; spinal and bulbar muscular atrophy;
KW dentatorubal-pallidoluysian atrophy; spinocerebellar ataxia type 1; SCA2;
KW SCA3; SCA4; SCA5; SCA6; SCA7; protein accumulation; intrabody.
XX
OS Synthetic.
XX
PN WO200106989-A2.
XX
PD 01-FEB-2001.
XX
PF 24-JUL-2000; 2000WO-US020131.
XX
PR 27-JUL-1999; 99US-0146047P.
PR 21-JUL-2000; 2000US-00620955.
XX
PA (HUST/) HUSTON J S.
PA (MESS/) MESSER A.
PA (LECE/) LECERF J.
XX
PI Huston JS, Messer A, Lecerf J;
XX
DR WPI; 2001-182700/18.
XX
PT Inhibiting intracellular polypeptide accumulation, useful for treating
PT neurological disorders, e.g. Alzheimer's disease, comprises contacting
PT the polypeptide with a specific intrabody.
XX
PS Disclosure; Page 99; 108pp; English.
XX
CC The present invention describes a method for inhibiting the formation of
CC aggregates of certain proteins, involving contacting the protein with a
CC binding molecule known as an intrabody. Proteins to be bound include
CC those associated with neurological disorders, and so the method can be
CC used in the prevention of diseases such as Alzheimer's, Parkinson's and
CC Huntington's diseases, prion disease, frontotemporal dementia,
CC amyotrophic lateral sclerosis, spinal and bulbar muscular atrophy,
CC dentatorubal-pallidoluysian atrophy, spinocerebellar ataxia type 1
CC (SCA1), SCA2, SCA3, SCA4, SCA5, SCA6 and SCA7
XX
SQ Sequence 66 AA;

Query Match 80.5%; Score 231; DB 4; Length 66;
Best Local Similarity 100.0%; Pred. No. 1.7e-19;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          5 GSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQLQP 51
              |||||||||||||||||||||||||||||||||||||||||||||||
Db         15 GSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQLQP 61
```

RESULT 3
AAB69614
ID AAB69614 standard; protein; 145 AA.
XX
AC AAB69614;
XX

DT 30-APR-2001 (first entry)
XX
DE Huntingtin accumulation inhibitor peptide GFP-HD-Q104.
XX
KW Neurological disorder; Huntington's disease; Alzheimer's disease;
KW Parkinson's disease; prion disease; frontotemporal dementia;
KW amyotrophic lateral sclerosis; spinal and bulbar muscular atrophy;
KW dentatorubal-pallidoluysian atrophy; spinocerebellar ataxia type 1; SCA2;
KW SCA3; SCA4; SCA5; SCA6; SCA7; protein accumulation; intrabody.
XX
OS Synthetic.
XX
PN WO200106989-A2.
XX
PD 01-FEB-2001.
XX
PF 24-JUL-2000; 2000WO-US020131.
XX
PR 27-JUL-1999; 99US-0146047P.
PR 21-JUL-2000; 2000US-00620955.
XX
PA (HUST/) HUSTON J S.
PA (MESS/) MESSER A.
PA (LECE/) LECERF J.
XX
PI Huston JS, Messer A, Lecerf J;
XX
DR WPI; 2001-182700/18.
XX
PT Inhibiting intracellular polypeptide accumulation, useful for treating
PT neurological disorders, e.g. Alzheimer's disease, comprises contacting
PT the polypeptide with a specific intrabody.
XX
PS Disclosure; Page 100; 108pp; English.
XX
CC The present invention describes a method for inhibiting the formation of
CC aggregates of certain proteins, involving contacting the protein with a
CC binding molecule known as an intrabody. Proteins to be bound include
CC those associated with neurological disorders, and so the method can be
CC used in the prevention of diseases such as Alzheimer's, Parkinson's and
CC Huntington's diseases, prion disease, frontotemporal dementia,
CC amyotrophic lateral sclerosis, spinal and bulbar muscular atrophy,
CC dentatorubal-pallidoluysian atrophy, spinocerebellar ataxia type 1
CC (SCA1), SCA2, SCA3, SCA4, SCA5, SCA6 and SCA7
XX
SQ Sequence 145 AA;

Query Match 76.0%; Score 218; DB 4; Length 145;
Best Local Similarity 97.8%; Pred. No. 1.2e-17;
Matches 45; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          5 GSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQ 50
              |||||||||||||||||||||||||||||||||||||||||||||| |
Db         15 GSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQ 60
```

RESULT 4

AAE26651
ID AAE26651 standard; protein; 63 AA.
XX
AC AAE26651;
XX
DT 13-DEC-2002 (first entry)
XX
DE Human huntington (htQ25) protein.
XX
KW Human; protein misfolding; Alzheimer's disease; AD; Parkinson's disease;
KW PD; Familial amyloid polyneuropathy; tauopathy; frontotemporal dementia;
KW Pick disease; lobar atrophy; trinucleotide disease; fragile-X syndrome;
KW Huntington's disease; spinocerebellar ataxia; SCA; myotonic dystrophy;
KW dentatorubral pallidoluysian atrophy; DRPLA; Creutzfeldt-Jacob disease;
KW CJD; prion disease; Gerstmann-Straussler-Scheinker disease; GSS; FFI;
KW fatal familia insomnia; mad cow disease; scrapie; kuru; anticonvulsant;
KW nootropic; neuroprotective; cerebroprotective; htQ25 protein.
XX
OS Homo sapiens.
XX
PN WO200265136-A2.
XX
PD 22-AUG-2002.
XX
PF 15-FEB-2002; 2002WO-US004632.
XX
PR 15-FEB-2001; 2001US-0269157P.
XX
PA (UYCH-) UNIV CHICAGO.
XX
PI Lindquist S, Krobitsch S, Outeiro T;
XX
DR WPI; 2002-667026/71.
DR N-PSDB; AAD44411.
XX
PT Screening for therapeutic agents for protein misfolding disease, by
PT contacting a yeast cell with compound, that expresses misfolded disease
PT protein, and with a toxicity inducing agent, and evaluating cell for
PT viability.
XX
PS Disclosure; Page 90; 93pp; English.
XX
CC The present invention relates to novel screening methods for identifying
CC therapeutic agents for diseases associated with protein misfolding. The
CC method involves contacting a yeast cell with a candidate compound, where
CC the yeast cell expresses a polypeptide comprising a misfolded disease
CC protein, contacting the yeast cell with a toxicity inducing agent and
CC evaluating the yeast cell for viability, where the viability indicates
CC the candidate compound is a candidate therapeutic agent. The method is
CC useful to screen for therapeutic agents for diseases associated with
CC protein misfolding such as Alzheimer's disease (AD), Parkinson's disease
CC (PD), Familial amyloid polyneuropathy, tauopathies (e.g. Pick disease,
CC lobar atrophy, frontotemporal dementia) or trinucleotide diseases (e.g.
CC Huntington's disease, spinocerebellar ataxia (SCA), fragile-X syndrome,
CC myotonic dystrophy, dentatorubral pallidoluysian atrophy (DRPLA) and
CC prion diseases (e.g. Creutzfeldt-Jacob disease (CJD), fatal familia
CC insomnia (FFI), Gerstmann-Straussler-Scheinker disease (GSS), mad cow

CC disease, scrapie and kuru). The method is useful for treating a patient
CC with Huntington's disease or Parkinson's disease. The present sequence is
CC human huntington (htQ25) protein. This sequence is used to illustrate the
CC method of the invention
XX
SQ Sequence 63 AA;

Query Match 72.5%; Score 208; DB 5; Length 63;
Best Local Similarity 95.6%; Pred. No. 7.6e-17;
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQLQP 51
              |||||||||||||||||||||||||||||||||||||||||  |
Db          1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQPPP 45
```

RESULT 5
AAB69607
ID AAB69607 standard; protein; 64 AA.
XX
AC AAB69607;
XX
DT 30-APR-2001 (first entry)
XX
DE Huntingtin accumulation inhibitor peptide HD-Q47-GFP.
XX
KW Neurological disorder; Huntington's disease; Alzheimer's disease;
KW Parkinson's disease; prion disease; frontotemporal dementia;
KW amyotrophic lateral sclerosis; spinal and bulbar muscular atrophy;
KW dentatorubal-pallidoluysian atrophy; spinocerebellar ataxia type 1; SCA2;
KW SCA3; SCA4; SCA5; SCA6; SCA7; protein accumulation; intrabody.
XX
OS Synthetic.
XX
PN WO200106989-A2.
XX
PD 01-FEB-2001.
XX
PF 24-JUL-2000; 2000WO-US020131.
XX
PR 27-JUL-1999; 99US-0146047P.
PR 21-JUL-2000; 2000US-00620955.
XX
PA (HUST/) HUSTON J S.
PA (MESS/) MESSER A.
PA (LECE/) LECERF J.
XX
PI Huston JS, Messer A, Lecerf J;
XX
DR WPI; 2001-182700/18.
XX
PT Inhibiting intracellular polypeptide accumulation, useful for treating
PT neurological disorders, e.g. Alzheimer's disease, comprises contacting
PT the polypeptide with a specific intrabody.
XX
PS Disclosure; Page 97; 108pp; English.
XX

```
CC   The present invention describes a method for inhibiting the formation of
CC   aggregates of certain proteins, involving contacting the protein with a
CC   binding molecule known as an intrabody. Proteins to be bound include
CC   those associated with neurological disorders, and so the method can be
CC   used in the prevention of diseases such as Alzheimer's, Parkinson's and
CC   Huntington's diseases, prion disease, frontotemporal dementia,
CC   amyotrophic lateral sclerosis, spinal and bulbar muscular atrophy,
CC   dentatorubal-pallidoluysian atrophy, spinocerebellar ataxia type 1
CC   (SCA1), SCA2, SCA3, SCA4, SCA5, SCA6 and SCA7
XX
SQ   Sequence 64 AA;

  Query Match             72.5%;  Score 208;  DB 4;  Length 64;
  Best Local Similarity   97.7%;  Pred. No. 7.7e-17;
  Matches   43;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQ 50
              ||||||||||||||||||||||||||||||||||||||||||| |
Db          1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQ 44


RESULT 6
AAW95073
ID   AAW95073 standard; protein; 86 AA.
XX
AC   AAW95073;
XX
DT   20-MAY-1999  (first entry)
XX
DE   GST-HD fusion protein GST-HD51DELP.
XX
KW   Amyloid-like fibril; protein aggregate; inhibitor; inclusion body;
KW   polyglutamine expansion; Huntington's disease; Alzheimer's disease; HD;
KW   Parkinson's disease; spinal; bulbar muscular atrophy; type II diabetes;
KW   systemic amyloidosis; spinocerebellar ataxia; kuru; familial insomnia;
KW   bovine spongiform encephalopathy; kuru; scrapie; GST-HD; fusion protein.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1
FT                   /note= "this residue is connected to a GST protein which
FT                   is not indicated in the sequence"
XX
PN   WO9906838-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998;   98WO-EP004810.
XX
PR   01-AUG-1997;   97EP-00113320.
XX
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI   Wanker E,  Lehrach H,  Scherzinger E,  Bates G;
XX
```

DR WPI; 1999-153955/13.
XX
PT Detecting amyloid-like fibrils or protein aggregates insoluble in
PT detergent or urea - from their retention on a filter, used for diagnosis,
PT particularly of diseases associated with polyglutamine expansion.
XX
PS Disclosure; Fig 8; 56pp; English.
XX
CC The invention relates to the detection of amyloid-like fibrils or protein
CC aggregates, insoluble in detergents or urea. The method comprises: (a)
CC applying material suspected of containing protein aggregates to a filter;
CC and (b) detecting retention of protein aggregates on the filter. This
CC method also helps to identify inhibitors of protein aggregates formation.
CC The method is particularly used to detect protein aggregates that are
CC indicative of disease, for assessing onset or progression of the
CC diseases. The inhibitors identified are potential therapeutic agents for
CC treating the diseases. Other applications include detection of inclusion
CC bodies in bacteria and to study kinetics of aggregate formation. Diseases
CC associated with polyglutamine expansion are particularly diagnosed, e.g.
CC Huntington's, Alzheimer's or Parkinson's diseases; spinal and bulbar
CC muscular atrophy; spinocerebellar ataxia; systemic amyloidosis; type II
CC diabetes; bovine spongiform encephalopathy; kuru; familial insomnia;
CC scrapie. The protein aggregates can now be detected simply, routinely and
CC rapidly, without requiring sophisticated equipment. The method can be
CC made quantitative, by analysing a series of dilutions, and can be
CC automated to allow many samples to be analysed on the same filter.
CC Sequences AAW95072-75 represent GST-HD fusion proteins
XX
SQ Sequence 86 AA;

Query Match 72.5%; Score 208; DB 2; Length 86;
Best Local Similarity 97.7%; Pred. No. 1e-16;
Matches 43; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQLQ 50
              ||||||||||||||||||||||||||||||||||||||||| |
Db          8 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQ 51
```

RESULT 7
AAW95078
ID AAW95078 standard; protein; 86 AA.
XX
AC AAW95078;
XX
DT 20-MAY-1999 (first entry)
XX
DE GST-HD fusion protein GST-HD51DELP.
XX
KW Fusion protein; amyloidogenic polypeptide; amyloid-like fibril; scrapie;
KW protein aggregate; Alzheimer's disease; CAG-repeat expansion; spinal;
KW Huntington's disease; bulbar muscular atrophy; spinocerebellar ataxia;
KW dentatorubral pallidoluysian atrophy; Creutzfeld-Jakob disease; enzyme;
KW GST-HD; HD.
XX
OS Synthetic.
OS Homo sapiens.

XX
FH Key Location/Qualifiers
FT Misc-difference 1
FT /note= "this residue is connected to a GST protein which
FT is not indicated in the sequence"
XX
PN WO9906545-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-EP004811.
XX
PR 01-AUG-1997; 97EP-00113306.
XX
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI Wanker E, Lehrach H, Scherzinger E, Bates G;
XX
DR WPI; 1999-153775/13.
XX
PT Composition containing fusion protein that includes amyloidogenic peptide
PT - able to self-assemble into fibrils or aggregates, used to detect and
PT monitor neuronal diseases, and also to screen for therapeutic inhibitors.
XX
PS Disclosure; Fig 8; 62pp; English.
XX
CC The invention relates to a composition comprising a fusion protein of (i)
CC (poly)peptide that increases solubility and/or prevents aggregation of
CC fusion protein, and (ii) amyloidogenic (poly)peptide that can self-
CC assemble into amyloid-like fibrils or protein aggregates. Host cells
CC transformed with a vector containing the nucleic acid encoding the fusion
CC protein are used for the recombinant expression of the fusion protein.
CC The composition is used to detect onset and progression of diseases
CC associated with fibrils/protein aggregates. It is potentially useful for
CC treatment of such diseases (e.g. Alzheimer's disease, scrapie or CAG-
CC repeat expansion conditions such as Huntington's disease (HD), spinal and
CC bulbar muscular atrophy, dentatorubral pallidoluysian atrophy,
CC spinocerebellar ataxia, Creutzfeld-Jakob disease). Assay methods based on
CC release of the amyloidogenic polypeptide from fusion protein have a
CC precise starting time for aggregate formation, allowing kinetic
CC measurements, and use of an enzyme for cleavage allows testing under
CC physiological conditions. Sequences AAW95077-80 represent GST-HD fusion
CC proteins
XX
SQ Sequence 86 AA;

Query Match 72.5%; Score 208; DB 2; Length 86;
Best Local Similarity 97.7%; Pred. No. 1e-16;
Matches 43; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQLQ 50
              ||||||||||||||||||||||||||||||||||||||||||| |
Db          8 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 51
```

RESULT 8
AAB69608

ID AAB69608 standard; protein; 89 AA.
XX
AC AAB69608;
XX
DT 30-APR-2001 (first entry)
XX
DE Huntingtin accumulation inhibitor peptide HD-Q72-GFP.
XX
KW Neurological disorder; Huntington's disease; Alzheimer's disease;
KW Parkinson's disease; prion disease; frontotemporal dementia;
KW amyotrophic lateral sclerosis; spinal and bulbar muscular atrophy;
KW dentatorubal-pallidoluysian atrophy; spinocerebellar ataxia type 1; SCA2;
KW SCA3; SCA4; SCA5; SCA6; SCA7; protein accumulation; intrabody.
XX
OS Synthetic.
XX
PN WO200106989-A2.
XX
PD 01-FEB-2001.
XX
PF 24-JUL-2000; 2000WO-US020131.
XX
PR 27-JUL-1999; 99US-0146047P.
PR 21-JUL-2000; 2000US-00620955.
XX
PA (HUST/) HUSTON J S.
PA (MESS/) MESSER A.
PA (LECE/) LECERF J.
XX
PI Huston JS, Messer A, Lecerf J;
XX
DR WPI; 2001-182700/18.
XX
PT Inhibiting intracellular polypeptide accumulation, useful for treating
PT neurological disorders, e.g. Alzheimer's disease, comprises contacting
PT the polypeptide with a specific intrabody.
XX
PS Disclosure; Page 97; 108pp; English.
XX
CC The present invention describes a method for inhibiting the formation of
CC aggregates of certain proteins, involving contacting the protein with a
CC binding molecule known as an intrabody. Proteins to be bound include
CC those associated with neurological disorders, and so the method can be
CC used in the prevention of diseases such as Alzheimer's, Parkinson's and
CC Huntington's diseases, prion disease, frontotemporal dementia,
CC amyotrophic lateral sclerosis, spinal and bulbar muscular atrophy,
CC dentatorubal-pallidoluysian atrophy, spinocerebellar ataxia type 1
CC (SCA1), SCA2, SCA3, SCA4, SCA5, SCA6 and SCA7
XX
SQ Sequence 89 AA;

  Query Match   72.5%; Score 208; DB 4; Length 89;
  Best Local Similarity   97.7%; Pred. No. 1.1e-16;
  Matches   43; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQLQ 50
              |||||||||||||||||||||||||||||||||||||||||| |

```
Db          1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 44


RESULT 9
AAW95075
ID   AAW95075 standard; protein; 94 AA.
XX
AC   AAW95075;
XX
DT   20-MAY-1999  (first entry)
XX
DE   GST-HD fusion protein GST-HD51DELPBio.
XX
KW   Amyloid-like fibril; protein aggregate; inhibitor; inclusion body;
KW   polyglutamine expansion; Huntington's disease; Alzheimer's disease; HD;
KW   Parkinson's disease; spinal; bulbar muscular atrophy; type II diabetes;
KW   systemic amyloidosis; spinocerebellar ataxia; kuru; familial insomnia;
KW   bovine spongiform encephalopathy; kuru; scrapie; GST-HD; fusion protein.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1
FT                   /note= "this residue is connected to a GST protein which
FT                   is not indicated in the sequence"
XX
PN   WO9906838-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998;   98WO-EP004810.
XX
PR   01-AUG-1997;   97EP-00113320.
XX
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI   Wanker E,  Lehrach H,  Scherzinger E,  Bates G;
XX
DR   WPI; 1999-153955/13.
XX
PT   Detecting amyloid-like fibrils or protein aggregates insoluble in
PT   detergent or urea - from their retention on a filter, used for diagnosis,
PT   particularly of diseases associated with polyglutamine expansion.
XX
PS   Disclosure; Fig 8; 56pp; English.
XX
CC   The invention relates to the detection of amyloid-like fibrils or protein
CC   aggregates, insoluble in detergents or urea. The method comprises: (a)
CC   applying material suspected of containing protein aggregates to a filter;
CC   and (b) detecting retention of protein aggregates on the filter. This
CC   method also helps to identify inhibitors of protein aggregates formation.
CC   The method is particularly used to detect protein aggregates that are
CC   indicative of disease, for assessing onset or progression of the
CC   diseases. The inhibitors identified are potential therapeutic agents for
CC   treating the diseases. Other applications include detection of inclusion
CC   bodies in bacteria and to study kinetics of aggregate formation. Diseases
```

```
CC   associated with polyglutamine expansion are particularly diagnosed, e.g.
CC   Huntington's, Alzheimer's or Parkinson's diseases; spinal and bulbar
CC   muscular atrophy; spinocerebellar ataxia; systemic amyloidosis; type II
CC   diabetes; bovine spongiform encephalopathy; kuru; familial insomnia;
CC   scrapie. The protein aggregates can now be detected simply, routinely and
CC   rapidly, without requiring sophisticated equipment. The method can be
CC   made quantitative, by analysing a series of dilutions, and can be
CC   automated to allow many samples to be analysed on the same filter.
CC   Sequences AAW95072-75 represent GST-HD fusion proteins
XX
SQ   Sequence 94 AA;

  Query Match             72.5%;  Score 208;  DB 2;  Length 94;
  Best Local Similarity   97.7%;  Pred. No. 1.1e-16;
  Matches   43;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQ 50
              |||||||||||||||||||||||||||||||||||||||||| |
Db          8 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQ 51


RESULT 10
AAW95080
ID   AAW95080 standard; protein; 94 AA.
XX
AC   AAW95080;
XX
DT   20-MAY-1999  (first entry)
XX
DE   GST-HD fusion protein GST-HD51DELPBio.
XX
KW   Fusion protein; amyloidogenic polypeptide; amyloid-like fibril; scrapie;
KW   protein aggregate; Alzheimer's disease; CAG-repeat expansion; spinal;
KW   Huntington's disease; bulbar muscular atrophy; spinocerebellar ataxia;
KW   dentatorubral pallidoluysian atrophy; Creutzfeld-Jakob disease; enzyme;
KW   GST-HD; HD.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1
FT                   /note= "this residue is connected to a GST protein which
FT                   is not indicated in the sequence"
XX
PN   WO9906545-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998;   98WO-EP004811.
XX
PR   01-AUG-1997;   97EP-00113306.
XX
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI   Wanker E,  Lehrach H,  Scherzinger E,  Bates G;
XX
```

DR WPI; 1999-153775/13.
XX
PT Composition containing fusion protein that includes amyloidogenic peptide
PT - able to self-assemble into fibrils or aggregates, used to detect and
PT monitor neuronal diseases, and also to screen for therapeutic inhibitors.
XX
PS Disclosure; Fig 8; 62pp; English.
XX
CC The invention relates to a composition comprising a fusion protein of (i)
CC (poly)peptide that increases solubility and/or prevents aggregation of
CC fusion protein, and (ii) amyloidogenic (poly)peptide that can self-
CC assemble into amyloid-like fibrils or protein aggregates. Host cells
CC transformed with a vector containing the nucleic acid encoding the fusion
CC protein are used for the recombinant expression of the fusion protein.
CC The composition is used to detect onset and progression of diseases
CC associated with fibrils/protein aggregates. It is potentially useful for
CC treatment of such diseases (e.g. Alzheimer's disease, scrapie or CAG-
CC repeat expansion conditions such as Huntington's disease (HD), spinal and
CC bulbar muscular atrophy, dentatorubral pallidoluysian atrophy,
CC spinocerebellar ataxia, Creutzfeld-Jakob disease). Assay methods based on
CC release of the amyloidogenic polypeptide from fusion protein have a
CC precise starting time for aggregate formation, allowing kinetic
CC measurements, and use of an enzyme for cleavage allows testing under
CC physiological conditions. Sequences AAW95077-80 represent GST-HD fusion
CC proteins
XX
SQ Sequence 94 AA;

Query Match 72.5%; Score 208; DB 2; Length 94;
Best Local Similarity 97.7%; Pred. No. 1.1e-16;
Matches 43; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQ 50
              ||||||||||||||||||||||||||||||||||||||||||| |
Db          8 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQ 51
```

RESULT 11
AAB69610
ID AAB69610 standard; protein; 98 AA.
XX
AC AAB69610;
XX
DT 30-APR-2001 (first entry)
XX
DE Huntingtin accumulation inhibitor peptide HD-Q47-Myc-HIS6.
XX
KW Neurological disorder; Huntington's disease; Alzheimer's disease;
KW Parkinson's disease; prion disease; frontotemporal dementia;
KW amyotrophic lateral sclerosis; spinal and bulbar muscular atrophy;
KW dentatorubal-pallidoluysian atrophy; spinocerebellar ataxia type 1; SCA2;
KW SCA3; SCA4; SCA5; SCA6; SCA7; protein accumulation; intrabody.
XX
OS Synthetic.
XX
PN WO200106989-A2.
XX

PD 01-FEB-2001.
XX
PF 24-JUL-2000; 2000WO-US020131.
XX
PR 27-JUL-1999; 99US-0146047P.
PR 21-JUL-2000; 2000US-00620955.
XX
PA (HUST/) HUSTON J S.
PA (MESS/) MESSER A.
PA (LECE/) LECERF J.
XX
PI Huston JS, Messer A, Lecerf J;
XX
DR WPI; 2001-182700/18.
XX
PT Inhibiting intracellular polypeptide accumulation, useful for treating
PT neurological disorders, e.g. Alzheimer's disease, comprises contacting
PT the polypeptide with a specific intrabody.
XX
PS Disclosure; Page 98; 108pp; English.
XX
CC The present invention describes a method for inhibiting the formation of
CC aggregates of certain proteins, involving contacting the protein with a
CC binding molecule known as an intrabody. Proteins to be bound include
CC those associated with neurological disorders, and so the method can be
CC used in the prevention of diseases such as Alzheimer's, Parkinson's and
CC Huntington's diseases, prion disease, frontotemporal dementia,
CC amyotrophic lateral sclerosis, spinal and bulbar muscular atrophy,
CC dentatorubal-pallidoluysian atrophy, spinocerebellar ataxia type 1
CC (SCA1), SCA2, SCA3, SCA4, SCA5, SCA6 and SCA7
XX
SQ Sequence 98 AA;

Query Match 72.5%; Score 208; DB 4; Length 98;
Best Local Similarity 97.7%; Pred. No. 1.2e-16;
Matches 43; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQLQ 50
              ||||||||||||||||||||||||||||||||||||||||||| |
Db          1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 44
```

RESULT 12
AAW95071
ID AAW95071 standard; protein; 108 AA.
XX
AC AAW95071;
XX
DT 20-MAY-1999 (first entry)
XX
DE Amino acid sequence of Huntington's gene exon 1 in GST-HD fusion protein.
XX
KW Amyloid-like fibril; protein aggregate; inhibitor; inclusion body;
KW polyglutamine expansion; Huntington's disease; Alzheimer's disease;
KW Parkinson's disease; spinal; bulbar muscular atrophy; type II diabetes;
KW systemic amyloidosis; spinocerebellar ataxia; kuru; familial insomnia;
KW bovine spongiform encephalopathy; kuru; scrapie; GST-HD.

XX
OS Synthetic.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 1
FT /note= "GST protein connected to the N-terminal"
FT Misc-difference 25
FT /note= "polyglutamine expansion that can comprise upto 51
FT glutamines"
XX
PN WO9906838-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-EP004810.
XX
PR 01-AUG-1997; 97EP-00113320.
XX
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI Wanker E, Lehrach H, Scherzinger E, Bates G;
XX
DR WPI; 1999-153955/13.
XX
PT Detecting amyloid-like fibrils or protein aggregates insoluble in
PT detergent or urea - from their retention on a filter, used for diagnosis,
PT particularly of diseases associated with polyglutamine expansion.
XX
PS Example 1; Fig 2; 56pp; English.
XX
CC The invention relates to the detection of amyloid-like fibrils or protein
CC aggregates, insoluble in detergents or urea. The method comprises: (a)
CC applying material suspected of containing protein aggregates to a filter;
CC and (b) detecting retention of protein aggregates on the filter. This
CC method also helps to identify inhibitors of protein aggregates formation.
CC The method is particularly used to detect protein aggregates that are
CC indicative of disease, for assessing onset or progression of the
CC diseases. The inhibitors identified are potential therapeutic agents for
CC treating the diseases. Other applications include detection of inclusion
CC bodies in bacteria and to study kinetics of aggregate formation. Diseases
CC associated with polyglutamine expansion are particularly diagnosed, e.g.
CC Huntington's, Alzheimer's or Parkinson's diseases; spinal and bulbar
CC muscular atrophy; spinocerebellar ataxia; systemic amyloidosis; type II
CC diabetes; bovine spongiform encephalopathy; kuru; familial insomnia;
CC scrapie. The protein aggregates can now be detected simply, routinely and
CC rapidly, without requiring sophisticated equipment. The method can be
CC made quantitative, by analysing a series of dilutions, and can be
CC automated to allow many samples to be analysed on the same filter. The
CC present sequence represents the Huntington's gene exon 1 translation
CC product which is connected to a GST protein to form a fusion protein. The
CC sequence of the GST protein is not indicated
XX
SQ Sequence 108 AA;

Query Match 72.5%; Score 208; DB 2; Length 108;
Best Local Similarity 97.7%; Pred. No. 1.3e-16;

```
  Matches   43;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQLQ 50
              ||||||||||||||||||||||||||||||||||||||||||| |
Db          8 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQ 51


RESULT 13
AAW95076
ID   AAW95076 standard; protein; 108 AA.
XX
AC   AAW95076;
XX
DT   20-MAY-1999  (first entry)
XX
DE   Amino acid sequence of Huntington's gene exon 1 in GST-HD fusion protein.
XX
KW   Fusion protein; amyloidogenic polypeptide; amyloid-like fibril; scrapie;
KW   protein aggregate; Alzheimer's disease; CAG-repeat expansion; spinal;
KW   Huntington's disease; bulbar muscular atrophy; spinocerebellar ataxia;
KW   dentatorubral pallidoluysian atrophy; Creutzfeld-Jakob disease; enzyme;
KW   GST-HD; HD.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1
FT                   /note= "GST protein connected to the N-terminal"
FT   Misc-difference 25
FT                   /note= "polyglutamine expansion that can comprise upto 51
FT                   glutamines"
XX
PN   WO9906545-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998;   98WO-EP004811.
XX
PR   01-AUG-1997;   97EP-00113306.
XX
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI   Wanker E,  Lehrach H,  Scherzinger E,  Bates G;
XX
DR   WPI; 1999-153775/13.
XX
PT   Composition containing fusion protein that includes amyloidogenic peptide
PT   - able to self-assemble into fibrils or aggregates, used to detect and
PT   monitor neuronal diseases, and also to screen for therapeutic inhibitors.
XX
PS   Example 1; Fig 2; 62pp; English.
XX
CC   The invention relates to a composition comprising a fusion protein of (i)
CC   (poly)peptide that increases solubility and/or prevents aggregation of
CC   fusion protein, and (ii) amyloidogenic (poly)peptide that can self-
CC   assemble into amyloid-like fibrils or protein aggregates. Host cells
```

CC transformed with a vector containing the nucleic acid encoding the fusion
CC protein are used for the recombinant expression of the fusion protein.
CC The composition is used to detect onset and progression of diseases
CC associated with fibrils/protein aggregates. It is potentially useful for
CC treatment of such diseases (e.g. Alzheimer's disease, scrapie or CAG-
CC repeat expansion conditions such as Huntington's disease (HD), spinal and
CC bulbar muscular atrophy, dentatorubral pallidoluysian atrophy,
CC spinocerebellar ataxia, Creutzfeld-Jakob disease). Assay methods based on
CC release of the amyloidogenic polypeptide from fusion protein have a
CC precise starting time for aggregate formation, allowing kinetic
CC measurements, and use of an enzyme for cleavage allows testing under
CC physiological conditions. The present sequence represents the
CC Huntington's gene exon 1 translation product which is connected to a GST
CC protein to form a fusion protein. The sequence of the GST protein is not
CC indicated
XX
SQ Sequence 108 AA;

Query Match 72.5%; Score 208; DB 2; Length 108;
Best Local Similarity 97.7%; Pred. No. 1.3e-16;
Matches 43; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQ 50
              ||||||||||||||||||||||||||||||||||||||||||| |
Db          8 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQ 51
```

RESULT 14
AAB69609
ID AAB69609 standard; protein; 121 AA.
XX
AC AAB69609;
XX
DT 30-APR-2001 (first entry)
XX
DE Huntingtin accumulation inhibitor peptide HD-Q104-GFP.
XX
KW Neurological disorder; Huntington's disease; Alzheimer's disease;
KW Parkinson's disease; prion disease; frontotemporal dementia;
KW amyotrophic lateral sclerosis; spinal and bulbar muscular atrophy;
KW dentatorubal-pallidoluysian atrophy; spinocerebellar ataxia type 1; SCA2;
KW SCA3; SCA4; SCA5; SCA6; SCA7; protein accumulation; intrabody.
XX
OS Synthetic.
XX
PN WO200106989-A2.
XX
PD 01-FEB-2001.
XX
PF 24-JUL-2000; 2000WO-US020131.
XX
PR 27-JUL-1999; 99US-0146047P.
PR 21-JUL-2000; 2000US-00620955.
XX
PA (HUST/) HUSTON J S.
PA (MESS/) MESSER A.
PA (LECE/) LECERF J.

XX
PI Huston JS, Messer A, Lecerf J;
XX
DR WPI; 2001-182700/18.
XX
PT Inhibiting intracellular polypeptide accumulation, useful for treating
PT neurological disorders, e.g. Alzheimer's disease, comprises contacting
PT the polypeptide with a specific intrabody.
XX
PS Disclosure; Page 98; 108pp; English.
XX
CC The present invention describes a method for inhibiting the formation of
CC aggregates of certain proteins, involving contacting the protein with a
CC binding molecule known as an intrabody. Proteins to be bound include
CC those associated with neurological disorders, and so the method can be
CC used in the prevention of diseases such as Alzheimer's, Parkinson's and
CC Huntington's diseases, prion disease, frontotemporal dementia,
CC amyotrophic lateral sclerosis, spinal and bulbar muscular atrophy,
CC dentatorubal-pallidoluysian atrophy, spinocerebellar ataxia type 1
CC (SCA1), SCA2, SCA3, SCA4, SCA5, SCA6 and SCA7
XX
SQ Sequence 121 AA;

Query Match 72.5%; Score 208; DB 4; Length 121;
Best Local Similarity 97.7%; Pred. No. 1.5e-16;
Matches 43; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQLQ 50
              ||||||||||||||||||||||||||||||||||||||||| |
Db          1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQ 44
```

RESULT 15
AAB69611
ID AAB69611 standard; protein; 123 AA.
XX
AC AAB69611;
XX
DT 30-APR-2001 (first entry)
XX
DE Huntingtin accumulation inhibitor peptide HD-Q72-Myc-HIS6.
XX
KW Neurological disorder; Huntington's disease; Alzheimer's disease;
KW Parkinson's disease; prion disease; frontotemporal dementia;
KW amyotrophic lateral sclerosis; spinal and bulbar muscular atrophy;
KW dentatorubal-pallidoluysian atrophy; spinocerebellar ataxia type 1; SCA2;
KW SCA3; SCA4; SCA5; SCA6; SCA7; protein accumulation; intrabody.
XX
OS Synthetic.
XX
PN WO200106989-A2.
XX
PD 01-FEB-2001.
XX
PF 24-JUL-2000; 2000WO-US020131.
XX
PR 27-JUL-1999; 99US-0146047P.

PR 21-JUL-2000; 2000US-00620955.
XX
PA (HUST/) HUSTON J S.
PA (MESS/) MESSER A.
PA (LECE/) LECERF J.
XX
PI Huston JS, Messer A, Lecerf J;
XX
DR WPI; 2001-182700/18.
XX
PT Inhibiting intracellular polypeptide accumulation, useful for treating
PT neurological disorders, e.g. Alzheimer's disease, comprises contacting
PT the polypeptide with a specific intrabody.
XX
PS Disclosure; Page 98-99; 108pp; English.
XX
CC The present invention describes a method for inhibiting the formation of
CC aggregates of certain proteins, involving contacting the protein with a
CC binding molecule known as an intrabody. Proteins to be bound include
CC those associated with neurological disorders, and so the method can be
CC used in the prevention of diseases such as Alzheimer's, Parkinson's and
CC Huntington's diseases, prion disease, frontotemporal dementia,
CC amyotrophic lateral sclerosis, spinal and bulbar muscular atrophy,
CC dentatorubal-pallidoluysian atrophy, spinocerebellar ataxia type 1
CC (SCA1), SCA2, SCA3, SCA4, SCA5, SCA6 and SCA7
XX
SQ Sequence 123 AA;

Query Match 72.5%; Score 208; DB 4; Length 123;
Best Local Similarity 97.7%; Pred. No. 1.5e-16;
Matches 43; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQLQ 50
              ||||||||||||||||||||||||||||||||||||||||||| |
Db          1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 44
```

Search completed: March 12, 2004, 15:38:30
Job time : 42.6471 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 12, 2004, 15:38:34 ; Search time 12.1471 Seconds
(without alignments)
250.755 Million cell updates/sec

Title: US-09-620-955B-10
Perfect score: 287
Sequence: 1 LVPRGSMATLEKLMKAFESL.....QQQQQQQQQLQPGSTRAAAS 59

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	196	68.3	513	3	US-09-041-886-28	Sequence 28, Appl
2	196	68.3	530	3	US-09-041-886-29	Sequence 29, Appl
3	196	68.3	552	3	US-09-041-886-30	Sequence 30, Appl
4	196	68.3	589	3	US-09-041-886-31	Sequence 31, Appl
5	196	68.3	3144	1	US-08-246-982A-6	Sequence 6, Appli
6	196	68.3	3144	1	US-08-453-265-6	Sequence 6, Appli
7	196	68.3	3144	2	US-08-457-273B-42	Sequence 42, Appl
8	196	68.3	3144	3	US-08-556-419-21	Sequence 21, Appl
9	196	68.3	3144	3	US-09-041-886-15	Sequence 15, Appl
10	143	49.8	1402	4	US-09-125-635-12	Sequence 12, Appl

11	142.5	49.7	528	4	US-09-086-663A-82	Sequence 82, Appl
12	142.5	49.7	548	4	US-09-086-663A-71	Sequence 71, Appl
13	142.5	49.7	596	4	US-09-086-663A-2	Sequence 2, Appli
14	142.5	49.7	596	4	US-09-086-663A-80	Sequence 80, Appl
15	140.5	49.0	513	3	US-09-100-193-3	Sequence 3, Appli
16	137	47.7	360	2	US-08-531-927B-2	Sequence 2, Appli
17	137	47.7	360	3	US-09-041-886-13	Sequence 13, Appl
18	136	47.4	2074	4	US-09-491-356C-9	Sequence 9, Appli
19	135	47.0	910	4	US-08-997-685A-2	Sequence 2, Appli
20	131	45.6	2023	4	US-09-491-356C-8	Sequence 8, Appli
21	130	45.3	1420	4	US-09-125-635-4	Sequence 4, Appli
22	129	44.9	521	4	US-09-086-663A-79	Sequence 79, Appl
23	129	44.9	521	4	US-09-086-663A-81	Sequence 81, Appl
24	129	44.9	1282	4	US-09-543-681A-5419	Sequence 5419, Ap
25	129	44.9	2703	1	US-08-185-432-19	Sequence 19, Appl
26	129	44.9	2703	4	US-08-899-232-4	Sequence 4, Appli
27	125	43.6	25	3	US-09-041-886-9	Sequence 9, Appli
28	123	42.9	816	2	US-08-267-803B-9	Sequence 9, Appli
29	123	42.9	816	3	US-09-041-886-17	Sequence 17, Appl
30	121.5	42.3	154	3	US-09-041-886-32	Sequence 32, Appl
31	121.5	42.3	918	3	US-09-041-886-11	Sequence 11, Appl
32	120	41.8	1088	4	US-09-233-857-13	Sequence 13, Appl
33	120	41.8	1099	4	US-09-442-100-2	Sequence 2, Appli
34	120	41.8	1099	4	US-08-939-106-2	Sequence 2, Appli
35	120	41.8	1099	4	US-09-442-102-2	Sequence 2, Appli
36	118	41.1	71	4	US-09-146-054-9	Sequence 9, Appli
37	118	41.1	71	4	US-09-664-977A-9	Sequence 9, Appli
38	118	41.1	795	1	US-07-716-827C-5	Sequence 5, Appli
39	118	41.1	1070	4	US-09-091-042A-2	Sequence 2, Appli
40	118	41.1	3119	1	US-08-246-982A-16	Sequence 16, Appl
41	118	41.1	3119	1	US-08-453-265-16	Sequence 16, Appl
42	117	40.8	428	1	US-08-190-802A-29	Sequence 29, Appl
43	117	40.8	428	3	US-08-477-346-29	Sequence 29, Appl
44	117	40.8	428	4	US-08-473-089-29	Sequence 29, Appl
45	117	40.8	428	4	US-08-487-072A-29	Sequence 29, Appl

ALIGNMENTS

RESULT 1
US-09-041-886-28
; Sequence 28, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabizadeh, Sharroz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/041,886
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 2626
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  513 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-041-886-28

  Query Match             68.3%;  Score 196;  DB 3;  Length 513;
  Best Local Similarity   91.1%;  Pred. No. 6.2e-16;
  Matches   41;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQP 51
              |||||||||||||||||||||||||||||||||||||||||    |
Db          1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQPPPPP 45


RESULT 2
US-09-041-886-29
; Sequence 29, Application US/09041886
; Patent No. 6235872
;  GENERAL INFORMATION:
;    APPLICANT:  Bredesen, Dale E.
;    APPLICANT:  Rabizadeh, Sharroz
;    TITLE OF INVENTION:  Proapoptotic Peptides, Dependence
;    TITLE OF INVENTION:  Polypeptides and Methods of Use
;    NUMBER OF SEQUENCES:  72
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/041,886
```

```
;       FILING DATE:
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Campbell, Cathryn A.
;       REGISTRATION NUMBER:  31,815
;       REFERENCE/DOCKET NUMBER:  P-LJ 2626
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (619) 535-9001
;       TELEFAX:  (619) 535-8949
;   INFORMATION FOR SEQ ID NO:  29:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  530 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
US-09-041-886-29

  Query Match             68.3%;  Score 196;  DB 3;  Length 530;
  Best Local Similarity   91.1%;  Pred. No. 6.5e-16;
  Matches   41;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQLQP 51
              |||||||||||||||||||||||||||||||||||||||||||    |
Db          1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQPPPPP 45


RESULT 3
US-09-041-886-30
; Sequence 30, Application US/09041886
; Patent No. 6235872
;  GENERAL INFORMATION:
;    APPLICANT:  Bredesen, Dale E.
;    APPLICANT:  Rabizadeh, Sharroz
;    TITLE OF INVENTION:  Proapoptotic Peptides, Dependence
;    TITLE OF INVENTION:  Polypeptides and Methods of Use
;    NUMBER OF SEQUENCES:  72
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/041,886
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 2626
;    TELECOMMUNICATION INFORMATION:
```

```
;       TELEPHONE:  (619) 535-9001
;       TELEFAX:  (619) 535-8949
;   INFORMATION FOR SEQ ID NO:  30:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  552 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
US-09-041-886-30

  Query Match             68.3%;  Score 196;  DB 3;  Length 552;
  Best Local Similarity   91.1%;  Pred. No. 6.8e-16;
  Matches   41;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQP 51
              |||||||||||||||||||||||||||||||||||||||||    |
Db          1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQPPPPP 45


RESULT 4
US-09-041-886-31
; Sequence 31, Application US/09041886
; Patent No. 6235872
;  GENERAL INFORMATION:
;    APPLICANT:  Bredesen, Dale E.
;    APPLICANT:  Rabizadeh, Sharroz
;    TITLE OF INVENTION:  Proapoptotic Peptides, Dependence
;    TITLE OF INVENTION:  Polypeptides and Methods of Use
;    NUMBER OF SEQUENCES:  72
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/041,886
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 2626
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  589 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  peptide
US-09-041-886-31

  Query Match             68.3%;  Score 196;  DB 3;  Length 589;
  Best Local Similarity   91.1%;  Pred. No. 7.3e-16;
  Matches   41;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQP 51
              |||||||||||||||||||||||||||||||||||||||||    |
Db          1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQPPPPP 45


RESULT 5
US-08-246-982A-6
; Sequence 6, Application US/08246982A
; Patent No. 5686288
;  GENERAL INFORMATION:
;    APPLICANT:  MacDonald, Marcy E.
;    APPLICANT:  Ambrose, Christine M.
;    APPLICANT:  Duyao, Mabel P.
;    APPLICANT:  Gusella, James F.
;    TITLE OF INVENTION:  Huntingtin DNA, Protein And Uses Thereof
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/246,982A
;      FILING DATE:  May 20, 1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldstein, Jorge, A.
;      REGISTRATION NUMBER:  29,021
;      REFERENCE/DOCKET NUMBER:  0609.3880002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3144 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-246-982A-6

  Query Match             68.3%;  Score 196;  DB 1;  Length 3144;
  Best Local Similarity   91.1%;  Pred. No. 4.6e-15;
  Matches   41;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;
```

```
Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQP 51
              ||||||||||||||||||||||||||||||||||||||||||    |
Db          1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQPPPPP 45
```

```
RESULT 6
US-08-453-265-6
; Sequence 6, Application US/08453265
; Patent No. 5693757
;  GENERAL INFORMATION:
;    APPLICANT:  MacDonald, Marcy E.
;    APPLICANT:  Ambrose, Christine M.
;    APPLICANT:  Duyao, Mabel P.
;    APPLICANT:  Gusella, James F.
;    TITLE OF INVENTION:  Huntingtin DNA, Protein And Uses Thereof
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/453,265
;      FILING DATE:  30-MAY-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ludwig, Steven R.
;      REGISTRATION NUMBER:  36,203
;      REFERENCE/DOCKET NUMBER:  0609.3880003
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3144 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-453-265-6

  Query Match             68.3%;  Score 196;  DB 1;  Length 3144;
  Best Local Similarity   91.1%;  Pred. No. 4.6e-15;
  Matches   41;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQP 51
              ||||||||||||||||||||||||||||||||||||||||||    |
Db          1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQPPPPP 45
```

```
RESULT 7
US-08-457-273B-42
; Sequence 42, Application US/08457273B
; Patent No. 5849995
;  GENERAL INFORMATION:
;    APPLICANT:  Hayden, Michael
;    APPLICANT:  Lin, Biaoyang
;    APPLICANT:  Nasir, Jamal
;    TITLE OF INVENTION:  Mouse Model for Huntington's Disease and
;    TITLE OF INVENTION:  Related DNA Sequences
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Virginia Bennett
;      STREET:  PO Box 37428
;      CITY:  Raleigh
;      STATE:  No. 5849995th Carolina
;      COUNTRY:  US
;      ZIP:  27627
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,273B
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bennett, Virginia C.
;      REGISTRATION NUMBER:  37,092
;      REFERENCE/DOCKET NUMBER:  3477-85A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-854-1400
;      TELEFAX:  919-854-1401
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3144 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-457-273B-42

  Query Match             68.3%;  Score 196;  DB 2;  Length 3144;
  Best Local Similarity   91.1%;  Pred. No. 4.6e-15;
  Matches   41;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQLQP 51
              |||||||||||||||||||||||||||||||||||||||||||    |
Db          1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQPPPPP 45


RESULT 8
US-08-556-419-21
; Sequence 21, Application US/08556419C
; Patent No. 6093549
; GENERAL INFORMATION:
```

```
;  APPLICANT: Ross, Christopher
;  APPLICANT: Li, Xiao-Jiang
;  APPLICANT: Li, Shi-Hua
;  APPLICANT: Sharp, Alan
;  APPLICANT: Lanahan, Anthony
;  APPLICANT: Worley, Paul
;  APPLICANT: Snyder, Solomon
;  TITLE OF INVENTION: Huntingtin-associated protein
;  FILE REFERENCE: 01107.52271
;  CURRENT APPLICATION NUMBER: US/08/556,419C
;  CURRENT FILING DATE: 1995-11-09
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 21
;   LENGTH: 3144
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-08-556-419-21

  Query Match             68.3%;  Score 196;  DB 3;  Length 3144;
  Best Local Similarity   91.1%;  Pred. No. 4.6e-15;
  Matches   41;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQP 51
              |||||||||||||||||||||||||||||||||||||||||    |
Db          1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQPPPP 45


RESULT 9
US-09-041-886-15
; Sequence 15, Application US/09041886
; Patent No. 6235872
;  GENERAL INFORMATION:
;    APPLICANT:  Bredesen, Dale E.
;    APPLICANT:  Rabizadeh, Sharroz
;    TITLE OF INVENTION:  Proapoptotic Peptides, Dependence
;    TITLE OF INVENTION:  Polypeptides and Methods of Use
;    NUMBER OF SEQUENCES:  72
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/041,886
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
```

```
;       REFERENCE/DOCKET NUMBER:  P-LJ 2626
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (619) 535-9001
;       TELEFAX:  (619) 535-8949
;   INFORMATION FOR SEQ ID NO:  15:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3144 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-09-041-886-15

  Query Match             68.3%;  Score 196;  DB 3;  Length 3144;
  Best Local Similarity   91.1%;  Pred. No. 4.6e-15;
  Matches   41;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQP 51
              |||||||||||||||||||||||||||||||||||||||||    |
Db          1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQPPPPP 45


RESULT 10
US-09-125-635-12
; Sequence 12, Application US/09125635
; Patent No. 6562589
; GENERAL INFORMATION:
;  APPLICANT:  THE UNITED STATES OF AMERICA represented by THE SE
;  TITLE OF INVENTION:  AIB1, A novel steriod receptor co-activator
;  FILE REFERENCE:  49944
;  CURRENT APPLICATION NUMBER: US/09/125,635
;  CURRENT FILING DATE:  1998-08-21
;  PRIOR APPLICATION NUMBER:  60/049,728
;  PRIOR FILING DATE:  1997-06-17
;  NUMBER OF SEQ ID NOS:  12
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 1402
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-125-635-12

  Query Match             49.8%;  Score 143;  DB 4;  Length 1402;
  Best Local Similarity   62.5%;  Pred. No. 4.7e-09;
  Matches   35;  Conservative    5;  Mismatches    8;  Indels    8;  Gaps    2;

Qy          2 VPR----GSMATL----EKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQL 49
              :||    ||: ||     :|  |  ||:  ||||||||||||||||||||||||:
Db        937 LPRPAMGGSVPTLPLRSNRLPGARPSLQQQQQQQQQQQQQQQQQQQQQQQQQQM 992


RESULT 11
US-09-086-663A-82
; Sequence 82, Application US/09086663A
; Patent No. 6518063
; GENERAL INFORMATION:
;  APPLICANT: DUCY, PATRICIA
;  APPLICANT:  KARSENTY, GERARD
```

```
;  TITLE OF INVENTION: OSF2/CBFA1 COMPOSITIONS AND METHODS OF USE
;  FILE REFERENCE: UTSC:525
;  CURRENT APPLICATION NUMBER: US/09/086,663A
;  CURRENT FILING DATE:  1998-05-29
;  PRIOR APPLICATION NUMBER: 60/080,189
;  PRIOR FILING DATE: 1998-03-24
;  PRIOR APPLICATION NUMBER: 60/048,430
;  PRIOR FILING DATE: 1997-05-29
;  NUMBER OF SEQ ID NOS: 83
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 82
;   LENGTH: 528
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-086-663A-82

  Query Match             49.7%;  Score 142.5;  DB 4;  Length 528;
  Best Local Similarity   58.2%;  Pred. No. 1.8e-09;
  Matches   32;  Conservative    7;  Mismatches   11;  Indels    5;  Gaps    1;

Qy          5 GSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQPGSTRAAAS 59
              | |: :  :: | :     ||||||||||||||||||||||||||| |  :  |||:
Db         37 GKMSDVSPVVAAQQ-----QQQQQQQQQQQQQQQQQQQQQQQQQQEAAAAAAAA 86


RESULT 12
US-09-086-663A-71
; Sequence 71, Application US/09086663A
; Patent No. 6518063
; GENERAL INFORMATION:
;  APPLICANT: DUCY, PATRICIA
;  APPLICANT:  KARSENTY, GERARD
;  TITLE OF INVENTION: OSF2/CBFA1 COMPOSITIONS AND METHODS OF USE
;  FILE REFERENCE: UTSC:525
;  CURRENT APPLICATION NUMBER: US/09/086,663A
;  CURRENT FILING DATE:  1998-05-29
;  PRIOR APPLICATION NUMBER: 60/080,189
;  PRIOR FILING DATE: 1998-03-24
;  PRIOR APPLICATION NUMBER: 60/048,430
;  PRIOR FILING DATE: 1997-05-29
;  NUMBER OF SEQ ID NOS: 83
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 71
;   LENGTH: 548
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-086-663A-71

  Query Match             49.7%;  Score 142.5;  DB 4;  Length 548;
  Best Local Similarity   58.2%;  Pred. No. 1.9e-09;
  Matches   32;  Conservative    7;  Mismatches   11;  Indels    5;  Gaps    1;

Qy          5 GSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQPGSTRAAAS 59
              | |: :  :: | :     ||||||||||||||||||||||||||| |  :  |||:
Db         37 GKMSDVSPVVAAQQ-----QQQQQQQQQQQQQQQQQQQQQQQQQQEAAAAAAAA 86
```

```
RESULT 13
US-09-086-663A-2
; Sequence 2, Application US/09086663A
; Patent No. 6518063
; GENERAL INFORMATION:
;  APPLICANT: DUCY, PATRICIA
;  APPLICANT:  KARSENTY, GERARD
;  TITLE OF INVENTION: OSF2/CBFA1 COMPOSITIONS AND METHODS OF USE
;  FILE REFERENCE: UTSC:525
;  CURRENT APPLICATION NUMBER: US/09/086,663A
;  CURRENT FILING DATE:  1998-05-29
;  PRIOR APPLICATION NUMBER: 60/080,189
;  PRIOR FILING DATE: 1998-03-24
;  PRIOR APPLICATION NUMBER: 60/048,430
;  PRIOR FILING DATE: 1997-05-29
;  NUMBER OF SEQ ID NOS: 83
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 596
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  Synthetic
;   OTHER INFORMATION: Peptide
US-09-086-663A-2

  Query Match             49.7%;  Score 142.5;  DB 4;  Length 596;
  Best Local Similarity   58.2%;  Pred. No. 2.1e-09;
  Matches   32;  Conservative    7;  Mismatches   11;  Indels    5;  Gaps    1;

Qy          5 GSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQPGSTRAAAS 59
              | |: :  :: | :     ||||||||||||||||||||||||| |  :  |||:
Db        105 GKMSDVSPVVAAQQ-----QQQQQQQQQQQQQQQQQQQQQQQQQQEAAAAAAAA 154


RESULT 14
US-09-086-663A-80
; Sequence 80, Application US/09086663A
; Patent No. 6518063
; GENERAL INFORMATION:
;  APPLICANT: DUCY, PATRICIA
;  APPLICANT:  KARSENTY, GERARD
;  TITLE OF INVENTION: OSF2/CBFA1 COMPOSITIONS AND METHODS OF USE
;  FILE REFERENCE: UTSC:525
;  CURRENT APPLICATION NUMBER: US/09/086,663A
;  CURRENT FILING DATE:  1998-05-29
;  PRIOR APPLICATION NUMBER: 60/080,189
;  PRIOR FILING DATE: 1998-03-24
;  PRIOR APPLICATION NUMBER: 60/048,430
;  PRIOR FILING DATE: 1997-05-29
;  NUMBER OF SEQ ID NOS: 83
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 80
;   LENGTH: 596
;   TYPE: PRT
```

```
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  Synthetic
;   OTHER INFORMATION: Peptide
US-09-086-663A-80

  Query Match             49.7%;  Score 142.5;  DB 4;  Length 596;
  Best Local Similarity   58.2%;  Pred. No. 2.1e-09;
  Matches   32;  Conservative    7;  Mismatches   11;  Indels    5;  Gaps    1;

Qy          5 GSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQPGSTRAAAS 59
              | |: :  :: | :     |||||||||||||||||||||||| |  :  |||:
Db        105 GKMSDVSPVVAAQQ-----QQQQQQQQQQQQQQQQQQQQQQQQQQEAAAAAAAA 154


RESULT 15
US-09-100-193-3
; Sequence 3, Application US/09100193
; Patent No. 6153729
;  GENERAL INFORMATION:
;    APPLICANT:  Gary S. Stein et al.
;    TITLE OF INVENTION:  NUCLEAR MATRIX TARGETING PEPTIDES AND USES THEREFORE
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  28 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/100,193
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/050,104
;      FILING DATE:  20-JUNE-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jane E. Remillard
;      REGISTRATION NUMBER:  38,872
;      REFERENCE/DOCKET NUMBER:  UMM-024
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)742-4214
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  513 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
US-09-100-193-3
```

Query Match 49.0%; Score 140.5; DB 3; Length 513;
Best Local Similarity 58.2%; Pred. No. 3.1e-09;
Matches 32; Conservative 6; Mismatches 10; Indels 7; Gaps 1;

```
Qy          5 GSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQLQPGSTRAAAS 59
              | |: :  :: |       |||||||||||||||||||||||| |  :  |||:
Db         23 GKMSDVSPVVAA-------QQQQQQQQQQQQQQQQQQQQQQQQQQQEAAAAAAA 70
```

Search completed: March 12, 2004, 15:42:40
Job time : 12.1471 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 12, 2004, 15:36:59 ; Search time 9.83333 Seconds
(without alignments)
577.149 Million cell updates/sec

Title: US-09-620-955B-10
Perfect score: 287
Sequence: 1 LVPRGSMATLEKLMKAFESL.....QQQQQQQQQLQPGSTRAAAS 59

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	196	68.3	3144	2	A46068	Huntington disease
2	147	51.2	1081	2	S66736	transcription acti
3	147	51.2	1761	2	T13675	hypothetical prote
4	145	50.5	339	1	TWHU2D	transcription init
5	145	50.5	1457	2	T14577	protein kinase Yak
6	141	49.1	966	2	S25365	CYC8 protein - yea
7	140.5	49.0	330	2	A35915	homeotic protein A
8	140.5	49.0	513	2	A48233	polyomavirus enhan
9	140	48.8	1572	2	S45251	SNF2alpha protein
10	139	48.4	2649	2	T51023	hypothetical prote
11	137	47.7	360	2	S50830	Machado-Joseph dis
12	136	47.4	139	2	A26892	Mopa box protein -
13	135	47.0	646	2	D82493	conserved hypothet

14	135	47.0	1154	2	S69206	regulator protein
15	134	46.7	1015	2	T13062	CLOCK protein - fr
16	134	46.7	1023	2	T13068	CLOCK protein - fr
17	134	46.7	1027	2	T13071	CLOCK protein - fr
18	133.5	46.5	445	1	S31224	transcription fact
19	133.5	46.5	445	1	A49447	transcription fact
20	133	46.3	796	2	JC7555	C14orf4 protein -
21	133	46.3	905	1	RGBYS5	regulatory protein
22	132.5	46.2	1586	2	S39580	HBRM protein - hum
23	132	46.0	356	2	S31574	hypothetical prote
24	131.5	45.8	1125	2	T14892	transcription fact
25	131	45.6	1969	2	T08875	histidine kinase h
26	130	45.3	623	2	A49840	segment polarity p
27	130	45.3	1424	2	T03851	thyroid hormone re
28	130	45.3	4957	2	T03455	ALR protein - huma
29	130	45.3	5262	2	T03454	ALR protein - huma
30	129	44.9	1090	2	A41696	regulatory protein
31	129	44.9	1905	2	T18267	multidrug resistan
32	129	44.9	2150	2	S71629	sensory transducti
33	129	44.9	2703	1	A24420	notch protein - fr
34	128	44.6	443	1	S29334	transcription fact
35	128	44.6	758	2	S54522	hypothetical prote
36	128	44.6	853	2	T46347	hypothetical prote
37	127	44.3	644	2	S39356	transcription fact
38	127	44.3	700	2	S09699	bib protein - frui
39	127	44.3	910	2	A34721	androgen receptor
40	127	44.3	911	2	B34721	androgen receptor
41	126.5	44.1	919	2	A39248	androgen receptor
42	126.5	44.1	1012	2	I53172	RAE-28 - mouse
43	126	43.9	1726	2	A39401	merozoite surface
44	126	43.9	3848	2	T17414	TipC protein - sli
45	125.5	43.7	539	2	S57972	hypothetical prote

ALIGNMENTS

RESULT 1
A46068
Huntington disease-associated protein - human
C;Species: Homo sapiens (man)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 08-Oct-1999
C;Accession: A46068; I54337
R;MacDonald, M.E.; Ambrose, C.M.; Duyao, M.P.; Myers, R.H.; Lin, C.; Srinidhi, L.; Barnes, G.; Taylor, S.A.; James, M.; Groot, N.; MacFarlane, H.; Jenkins, B.; Anderson, M.A.; Wexler, N.S.; Gusella, J.F.; Bates, G.P.; Baxendale, S.; Hummerich, H.; Kirby, S.; North, M.; Youngman, S.; Mott, R.; Zehetner, G.; Sedlacek, Z.; Poustka, A.; Frischauf, A.M.; Buckler, A.J.; Church, D.; Doucette-Stamm, L.; O'Donovan, M.C.; Riba-Ramirez, L.; Shah, M.; Stanton, V.P.; Strobel, S.A.; Draths, K.M.
Cell 72, 971-983, 1993
A;Authors: Wales, J.L.; Dervan, P.; Housman, D.E.; Altherr, M.; Shiang, R.; Thompson, L.; Fielder, T.; Wasmuth, J.J.; Tagle, D.; Valdes, J.; Elmer, L.; Allard, M.; Castilla, L.; Swaroop, M.; Blanchard, K.; Collins, F.S.; Snell, R.; Holloway, T.; Gillespie, K.; Datson, N.; Shaw, D.; Harper, P.S.
A;Title: A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes.

A;Reference number: A46068; MUID:93208892; PMID:8458085
A;Accession: A46068
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-3144 <MAC>
A;Cross-references: GB:L12392
R;Lin, B.; Rommens, J.M.; Graham, R.K.; Kalchman, M.; MacDonald, H.; Nasir, J.; Delaney, A.; Goldberg, Y.P.; Hayden, M.R.
Hum. Mol. Genet. 2, 1541-1545, 1993
A;Title: Differential 3' polyadenylation of the Huntington disease gene results in two mRNA species with variable tissue expression.
A;Reference number: I54337; MUID:94093536; PMID:7903579
A;Accession: I54337
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 2563-3144 <RES>
A;Cross-references: GB:L20431; NID:g398028; PIDN:AAA52702.1; PID:g398029
C;Genetics:
A;Gene: GDB:HD
A;Cross-references: GDB:119307; OMIM:143100
A;Map position: 4p16.3-4p16.3

  Query Match 68.3%; Score 196; DB 2; Length 3144;
  Best Local Similarity 91.1%; Pred. No. 5.8e-11;
  Matches 41; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQP 51
              |||||||||||||||||||||||||||||||||||||||    |
Db          1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQPPPPP 45
```

RESULT 2
S66736
transcription activator GAL11 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein O1280; protein YOL051w
C;Species: Saccharomyces cerevisiae
C;Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 23-Mar-2001
C;Accession: S66736; S66743; S59300; S61730; S45695; A31565
R;Ansorge, W.; Benes, V.; Rechmann, S.; Schwager, C.; Teodoru, C.; Voss, H.; Wiemann, S.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S66723
A;Accession: S66736
A;Molecule type: DNA
A;Residues: 1-1081 <ANS>
A;Cross-references: EMBL:Z74793; NID:g1419855; PID:e252273; PID:g1419856; MIPS:YOL051w
A;Experimental source: strain S288C
R;Feldmann, H.; Mannhaupt, G.; Vetter, I.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S66743
A;Accession: S66743
A;Molecule type: DNA
A;Residues: 1-351 <FEL>
A;Cross-references: EMBL:Z74793; MIPS:YOL051w
A;Experimental source: strain S288C
R;Mannhaupt, G.; Vetter, I.; Schwarzlose, C.; Mitzel, S.; Feldmann, H.

submitted to the EMBL Data Library, August 1995
A;Description: Analysis of a 26kb region on the left arm of yeast chromosome XV.
A;Reference number: S59285
A;Accession: S59300
A;Molecule type: DNA
A;Residues: 1-351 <FEW>
A;Cross-references: EMBL:X91067; NID:g984177; PID:g984193
R;Mannhaupt, G.; Vetter, I.; Schwarzlose, C.; Mitzel, S.; Feldmann, H.
Yeast 12, 67-76, 1996
A;Title: Analysis of a 26 kb region on the left arm of yeast chromosome XV.
A;Reference number: S61715; MUID:96381248; PMID:8789261
A;Accession: S61730
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-351 <MAN>
A;Cross-references: EMBL:X91067; NID:g984177; PIDN:CAA62537.1; PID:g984193
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August
1995
R;Suzuki, Y.; Nogi, Y.; Abe, A.; Fukasawa, T.
Mol. Cell. Biol. 12, 4806, 1992
A;Reference number: S45695; MUID:93024425; PMID:1406662
A;Contents: erratum
A;Accession: S45695
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-170,'T',172-301,'Q',303-498,'T',500-750,'Q',752-1081 <SUZ1>
A;Cross-references: EMBL:M22481; NID:g171549; PID:g171550
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August
1992
A;Note: this is a revision to the sequence from reference A31565
R;Suzuki, Y.; Nogi, Y.; Abe, A.; Fukasawa, T.
Mol. Cell. Biol. 8, 4991-4999, 1988
A;Title: GAL11 protein, an auxiliary transcription activator for genes encoding
galactose-metabolizing enzymes in Saccharomyces cerevisiae.
A;Reference number: A31565; MUID:89096873; PMID:3062377
A;Accession: A31565
A;Molecule type: DNA
A;Residues: 118-1081 <SUZ2>
A;Cross-references: EMBL:M22481
A;Note: this sequence has been revised in reference S45695
C;Genetics:
A;Gene: SGD:GAL11
A;Cross-references: SGD:S0005411; MIPS:YOL051w
A;Map position: 15L
C;Keywords: transcription regulation

  Query Match   51.2%; Score 147; DB 2; Length 1081;
  Best Local Similarity   51.7%; Pred. No. 1.1e-06;
  Matches   30; Conservative   11; Mismatches   13; Indels   4; Gaps   1;

Qy   6 SMATLEKLMKAFESLKSFQ----QQQQQQQQQQQQQQQQQQQQQQQQLQPGSTRAAAS 59
              ::|| : : :: : :: | |||||||||||||||||||||| : | || |:
Db   651 NIATQQNMQQSLQQMQHLQQLKMQQQQQQQQQQQQQQQQQQQQQQQQHIYPSSTPGVAN 708

RESULT 3
T13675

hypothetical protein EG0002.3 - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13675
R;Bolshakov, V.; Borkova, D.; Minana, B.; Kafatos, F.
submitted to the EMBL Data Library, September 1998
A;Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A;Reference number: Z17698
A;Accession: T13675
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1761 <BOL>
A;Cross-references: EMBL:AL031130; NID:e1316407; PID:e1316410; PIDN:CAA20016.1
C;Genetics:
A;Cross-references: FlyBase:FBgn0025376
A;Introns: 143/3; 237/3; 280/3
A;Note: EG:EG0002.3

  Query Match   51.2%; Score 147; DB 2; Length 1761;
  Best Local Similarity   60.4%; Pred. No. 1.7e-06;
  Matches   29; Conservative   8; Mismatches   11; Indels   0; Gaps   0;

Qy   3 PRGSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQ 50
              | |: | ::: :: : : ||||||||||||||||||||||||| |
Db   1474 PAGATADMQRYVQRMQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 1521

RESULT 4
TWHU2D
transcription initiation factor IID - human
N;Alternate names: TATA-binding protein
C;Species: Homo sapiens (man)
C;Date: 20-Jul-1990 #sequence_revision 19-May-1995 #text_change 18-Feb-2000
C;Accession: A34830; A34831; S10944; I60128
R;Peterson, M.G.; Tanese, N.; Pugh, B.F.; Tjian, R.
Science 248, 1625-1630, 1990
A;Title: Functional domains and upstream activation properties of cloned human
TATA binding protein.
A;Reference number: A34830; MUID:90302006; PMID:2363050
A;Accession: A34830
A;Molecule type: mRNA
A;Residues: 1-339 <PET>
A;Cross-references: GB:M55654; NID:g339491; PIDN:AAA36731.1; PID:g339492
R;Kao, C.C.; Lieberman, P.M.; Schmidt, M.C.; Zhou, Q.; Pei, R.; Berk, A.J.
Science 248, 1646-1649, 1990
A;Title: Cloning of a transcriptionally active human TATA binding factor.
A;Reference number: A34831; MUID:90302010; PMID:2194289
A;Accession: A34831
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-17,'N',19-186,'R',188-339 <KAO>
R;Hoffmann, A.; Sinn, E.; Yamamoto, T.; Wang, J.; Roy, A.; Horikoshi, M.;
Roeder, R.G.
Nature 346, 387-390, 1990
A;Title: Highly conserved core domain and unique N terminus with presumptive
regulatory motifs in a human TATA factor (TFIID).
A;Reference number: S10944; MUID:90326195; PMID:2374612

A;Accession: S10944
A;Molecule type: mRNA
A;Residues: 1-91,96-339 <HOF>
A;Cross-references: EMBL:X54993; NID:g37065; PIDN:CAA38736.1; PID:g37066
R;Kao, C.; Lieberman, P.; Schmidt, M.; Zhou, Q.; Pei, R.; Berk, A.J.
Science 248, 1626, 1990
A;Title: Cloning of the human TATA binding factor: Expression of a
transcriptionally active TFIID protein.
A;Reference number: I60128
A;Accession: I60128
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-186,'R',188-299,'MIKPR',300-339 <RES>
A;Cross-references: GB:M34960; NID:g339493; PID:g339494
C;Genetics:
A;Gene: GDB:TBP; GTF2D1
A;Cross-references: GDB:138768; OMIM:600075
A;Map position: 6q27-6q27
C;Superfamily: human transcription initiation factor IID
C;Keywords: alternative splicing; DNA binding; nucleus; transcription initiation
F;55-95/Region: glutamine-rich

  Query Match   50.5%; Score 145; DB 1; Length 339;
  Best Local Similarity   59.3%; Pred. No. 5.9e-07;
  Matches   32; Conservative   7; Mismatches   15; Indels   0; Gaps   0;

Qy   6 SMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQPGSTRAAAS 59
              |:: ||: : : : |||||||||||||||||||||||| | |||:
Db   48 SLSILEEQQRQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQAVAAAA 101

RESULT 5
T14577
protein kinase YakA (EC 2.7.1.-) - slime mold (Dictyostelium discoideum)
C;Species: Dictyostelium discoideum
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T14577
R;Kuspa, A.; Lu, S.; Souza, G.M.
submitted to the EMBL Data Library, January 1998
A;Description: YakA, a protein kinase required for the growth to development
transition in Dictyostelium.
A;Reference number: Z18146
A;Accession: T14577
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1457 <KUS>
A;Cross-references: EMBL:AF045453; NID:g2854116; PID:g2854117; PIDN:AAC02554.1
C;Genetics:
A;Gene: yakA
C;Keywords: ATP; phosphoprotein; phosphotransferase; serine/threonine-specific
protein kinase

  Query Match   50.5%; Score 145; DB 2; Length 1457;
  Best Local Similarity   59.2%; Pred. No. 2.2e-06;
  Matches   29; Conservative   7; Mismatches   13; Indels   0; Gaps   0;

Qy   2 VPRGSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQ 50

```
                :|: ||    :::   :  :  |||||||||||||||||||||||| |
Db          575 IPQHSMLNGNQILNQHQLFQQLQQQQQQQQQQQQQQQQQQQQQQQQQ 623
```

RESULT 6
S25365
CYC8 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: glucose repression mediator; protein YBR0908; protein YBR112c; SSN6 protein
C;Species: Saccharomyces cerevisiae
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 11-Jan-2000
C;Accession: S25365; S48277; S45980; S25404; S25405; A30906; S44692
R;Mannhaupt, G.; Stucka, R.; Ehnle, S.; Vetter, I.; Feldmann, H.
Yeast 8, 397-408, 1992
A;Title: Molecular analysis of yeast chromosome II between CMD1 and LYS2: the excision repair gene RAD16 located in this region belongs to a novel group of double-finger proteins.
A;Reference number: S25364; MUID:92327848; PMID:1626431
A;Accession: S25365
A;Molecule type: DNA
A;Residues: 1-966 <MAN>
A;Cross-references: EMBL:X66247; NID:g3548; PIDN:CAA46973.1; PID:g3550
R;Mannhaupt, G.; Stucka, R.; Ehnle, S.; Vetter, I.; Feldmann, H.
Yeast 10, 1363-1381, 1994
A;Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.
A;Reference number: S48255; MUID:95208357; PMID:7900426
A;Accession: S48277
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-966 <MAW>
A;Cross-references: EMBL:X78993; NID:g476045; PIDN:CAA55615.1; PID:g476068
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1994
R;Feldmann, H.; Mannhaupt, G.; Schwarzlose, C.; Vetter, I.
submitted to the Protein Sequence Database, August 1994
A;Reference number: S45927
A;Accession: S45980
A;Molecule type: DNA
A;Residues: 1-966 <FE2>
A;Cross-references: EMBL:Z35981; NID:g536449; PIDN:CAA85069.1; PID:g536450; MIPS:YBR112c
R;Schultz, J.; Carlson, M.
Mol. Cell. Biol. 7, 3637-3645, 1987
A;Title: Molecular analysis of SSN6, a gene functionally related to the SNF1 protein kinase of Saccharomyces cerevisiae.
A;Reference number: S25404; MUID:88065502; PMID:3316983
A;Accession: S25404
A;Molecule type: DNA
A;Residues: 1-546,'K',548-966 <SCH>
A;Cross-references: EMBL:M17826; NID:g172725; PIDN:AAA35103.1; PID:g172726
R;Trumbly, R.J.
Gene 73, 97-111, 1988
A;Title: Cloning and characterization of the CYC8 gene mediating glucose repression in yeast.
A;Reference number: S25405; MUID:89211964; PMID:2854095
A;Accession: S25405
A;Molecule type: DNA

A;Residues: 1-546,'K',548-966 <TRU>
A;Cross-references: EMBL:M23440; NID:g171349; PIDN:AAA34545.1; PID:g171350
C;Genetics:
A;Gene: SGD:CYC8; SSN6; CRT8
A;Cross-references: SGD:S0000316; MIPS:YBR112c
A;Map position: 2R
C;Function:
A;Description: required for complete derepression of ICL1; required for repression of SUC2 at high glucose levels and for induction of SUC2 at low glucose levels
C;Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repeat homology
C;Keywords: nucleus; transcription regulation
F;224-257/Domain: tetratricopeptide repeat homology <TT1>
F;262-295/Domain: tetratricopeptide repeat homology <TT2>
F;296-329/Domain: tetratricopeptide repeat homology <TT3>
F;330-363/Domain: tetratricopeptide repeat homology <TT4>
F;365-398/Domain: tetratricopeptide repeat homology <TT5>

  Query Match 49.1%; Score 141; DB 2; Length 966;
  Best Local Similarity 100.0%; Pred. No. 3.6e-06;
  Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 24 QQQQQQQQQQQQQQQQQQQQQQQQQLQP 51
       ||||||||||||||||||||||||||||
Db 563 QQQQQQQQQQQQQQQQQQQQQQQQQLQP 590

RESULT 7
A35915
homeotic protein Abdominal-A - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 08-Mar-1991 #sequence_revision 08-Mar-1991 #text_change 24-Sep-1999
C;Accession: A35915
R;Karch, F.; Bender, W.; Weiffenbach, B.
Genes Dev. 4, 1573-1587, 1990
A;Title: abdA expression in Drosophila embryos.
A;Reference number: A35915; MUID:91071585; PMID:1979297
A;Accession: A35915
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-330 <KAR>
A;Cross-references: GB:X54453; NID:g7522; PIDN:CAA38321.1; PID:g7523
C;Genetics:
A;Gene: FlyBase:abd-A
A;Cross-references: FlyBase:FBgn0000014
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;139-195/Domain: homeobox homology <HOX>
F;227-248/Region: glutamine-rich

  Query Match 49.0%; Score 140.5; DB 2; Length 330;
  Best Local Similarity 70.5%; Pred. No. 1.5e-06;
  Matches 31; Conservative 4; Mismatches 6; Indels 3; Gaps 1;

Qy 11 EKLMKAFESLKSFQQQ---QQQQQQQQQQQQQQQQQQQQLQP 51
       :: ||| |::|| || |||||||||||||||| |||| ||

```
Db        214 QEKMKAQETMKSAQQNKQVQQQQQQQQQQQQQQQQQHQQQQQQP 257
```

RESULT 8
A48233
polyomavirus enhancer-binding protein 2 alpha chain type 1 - mouse
N;Alternate names: PEA2 alpha chain type 1; PEA2 alpha chain type 2; PEBP2 alpha chain type 1; PEBP2 alpha chain type 2
C;Species: Mus musculus (house mouse)
C;Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 01-Dec-2000
C;Accession: A48233; B48233
R;Ogawa, E.; Maruyama, M.; Kagoshima, H.; Inuzuka, M.; Lu, J.; Satake, M.; Shigesada, K.; Ito, Y.
Proc. Natl. Acad. Sci. U.S.A. 90, 6859-6863, 1993
A;Title: PEBP2/PEA2 represents a family of transcription factors homologous to the products of the Drosophila runt gene and the human AML1 gene.
A;Reference number: A48233; MUID:93342088; PMID:8341710
A;Accession: A48233
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-513 <OGA>
A;Cross-references: GB:D14636; NID:g391766; PIDN:BAA03485.1; PID:d1003996; PID:g391767
A;Accession: B48233
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-304,'L',306 <OG2>
A;Cross-references: GB:D14637; NID:g391768; PIDN:BAA03486.1; PID:g391769
C;Genetics:
A;Gene: PEBP2alphaA
C;Superfamily: transcription factor CBF alpha 2
C;Keywords: alternative splicing; DNA binding; T-cell; transcription factor; transcription regulation

Query Match 49.0%; Score 140.5; DB 2; Length 513;
Best Local Similarity 58.2%; Pred. No. 2.3e-06;
Matches 32; Conservative 6; Mismatches 10; Indels 7; Gaps 1;

```
Qy          5 GSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQLQPGSTRAAAS 59
              | |: :  :: |       ||||||||||||||||||||||| |  :  |||:
Db         23 GKMSDVSPVVAA-------QQQQQQQQQQQQQQQQQQQQQQQQQQQEAAAAAAA 70
```

RESULT 9
S45251
SNF2alpha protein - human
C;Species: Homo sapiens (man)
C;Date: 10-Dec-1994 #sequence_revision 10-Nov-1995 #text_change 02-Aug-2002
C;Accession: S45251
R;Chiba, H.; Muramatsu, M.; Nomoto, A.; Kato, H.
Nucleic Acids Res. 22, 1815-1820, 1994
A;Title: Two human homologues of Saccharomyces cerevisiae SWI2/SNF2 and Drosophila brahma are transcriptional coactivators cooperating with the estrogen receptor and the retinoic acid receptor.
A;Reference number: S45251; MUID:94268902; PMID:8208605
A;Accession: S45251
A;Status: preliminary

A;Molecule type: mRNA
A;Residues: 1-1572 <CHI>
A;Cross-references: GB:D26155; NID:g505086; PIDN:BAA05142.1; PID:d1005684;
PID:g987661
C;Superfamily: human SNF2alpha protein; bromodomain homology
F;1409-1464/Domain: bromodomain homology <BRO>

  Query Match   48.8%; Score 140; DB 2; Length 1572;
  Best Local Similarity   67.4%; Pred. No. 7e-06;
  Matches   29; Conservative   4; Mismatches   10; Indels   0; Gaps   0;

```
Qy          9 TLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQP 51
              ||:  ::   :|   |||||||||||||||||||||| | ||
Db        201 TLQLAVQGKRTLPGLQQQQQQQQQQQQQQQQQQQQQQPQQQP 243
```

RESULT 10
T51023
hypothetical protein B7F21.40 [imported] - Neurospora crassa
C;Species: Neurospora crassa
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C;Accession: T51023
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.;
Nyakatura, G.; Mewes, H.W.; Mannhaupt, G.
submitted to the Protein Sequence Database, July 2000
A;Reference number: Z25286
A;Accession: T51023
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2649 <SCH>
A;Cross-references: EMBL:AL389901; GSPDB:GN00116; NCSP:B7F21.40
A;Experimental source: BAC clone B7F21; strain OR74A
C;Genetics:
A;Gene: NCSP:B7F21.40
A;Map position: 6
A;Introns: 1619/3; 2584/1

  Query Match   48.4%; Score 139; DB 2; Length 2649;
  Best Local Similarity   61.2%; Pred. No. 1.4e-05;
  Matches   30; Conservative   5; Mismatches   14; Indels   0; Gaps   0;

```
Qy          4 RGSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQPG 52
              |   :: ::   |   |:  |||||||||||||||||||||||| | |
Db       2207 REEWSSTQQGQAAVSGLQQRQQQQQQQQQQQQQQQQQQQQQQQQQG 2255
```

RESULT 11
S50830
Machado-Joseph disease MJD1a protein - human
C;Species: Homo sapiens (man)
C;Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 28-May-1999
C;Accession: S50830
R;Kawaguchi, Y.; Okamoto, T.; Taniwaki, M.; Aizawa, M.; Inoue, M.; Katayama, S.;
Kawakami, H.; Nakamura, S.; Nishimura, M.; Akiguchi, I.; Kimura, J.; Narumiya,
S.; Kakizuka, A.
Nature Genet. 8, 221-228, 1994

A;Title: CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1.
A;Reference number: S50830; MUID:95179166; PMID:7874163
A;Accession: S50830
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-360 <KAW>
A;Cross-references: GB:S75313; NID:g833927; PIDN:AAB33571.1; PID:g833928

  Query Match 47.7%; Score 137; DB 2; Length 360;
  Best Local Similarity 68.2%; Pred. No. 3.6e-06;
  Matches 30; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

Qy 5 GSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQ 48
Db 273 GTNLTSEELRKRREAYFEKQQQKQQQQQQQQQQQQQQQQQQQQ 316

RESULT 12
A26892
Mopa box protein - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 05-Nov-1999
C;Accession: A26892
R;Duboule, D.; Haenlin, M.; Galliot, B.; Mohier, E.
Mol. Cell. Biol. 7, 2003-2006, 1987
A;Title: DNA sequences homologous to the Drosophila opa repeat are present in murine mRNAs that are differentially expressed in fetuses and adult tissues.
A;Reference number: A26892; MUID:87257908; PMID:2885744
A;Accession: A26892
A;Molecule type: mRNA
A;Residues: 1-139 <DUB>
A;Cross-references: GB:M16362; NID:g200142; PIDN:AAA39860.1; PID:g387503

  Query Match 47.4%; Score 136; DB 2; Length 139;
  Best Local Similarity 77.8%; Pred. No. 1.9e-06;
  Matches 28; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 24 QQQQQQQQQQQQQQQQQQQQQQQQQLQPGSTRAAAS 59
Db 52 QQQQQQQQQQQQQQQQQQQQQQQQQPHQQQQQAA 87

RESULT 13
D82493
conserved hypothetical protein VCA0171 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: D82493
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Umayam, L.A.; Gill, S.R.; Nelson, K.E.; Read, T.D.; Tettelin, H.; Richardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.; McDonald, L.; Utterback, T.; Fleishmann, R.D.; Nierman, W.C.; White, O.; Salzberg, S.L.; Smith, H.O.; Colwell, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000

A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833; PMID:10952301
A;Accession: D82493
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-646 <HEI>
A;Cross-references: GB:AE004357; GB:AE003853; NID:g9657547; PIDN:AAF96084.1; GSPDB:GN00127; TIGR:VCA0171
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VCA0171
A;Map position: 2

  Query Match 47.0%; Score 135; DB 2; Length 646;
  Best Local Similarity 73.7%; Pred. No. 9.5e-06;
  Matches 28; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 13 LMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQ 50
       ::|| : : |||||||||||||||||||||||| |
Db 439 VVKAAQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 476

RESULT 14
S69206
regulator protein white collar 1 - Neurospora crassa
C;Species: Neurospora crassa
C;Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 11-Jan-2002
C;Accession: S69206
R;Ballario, P.; Vittorioso, P.; Magrelli, A.; Talora, C.; Cabibbo, A.; Macino, G.
EMBO J. 15, 1650-1657, 1996
A;Title: White collar-1, a central regulator of blue light responses in Neurospora, is a zinc finger protein.
A;Reference number: S69206; MUID:96203083; PMID:8612589
A;Accession: S69206
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1154 <BAL>
A;Cross-references: EMBL:X94300; NID:g1279576; PID:g1480115
C;Genetics:
A;Introns: 967/3
C;Superfamily: GATA-type zinc finger homology
C;Keywords: zinc finger
F;932-991/Domain: GATA-type zinc finger homology <GZF>

  Query Match 47.0%; Score 135; DB 2; Length 1154;
  Best Local Similarity 65.2%; Pred. No. 1.6e-05;
  Matches 30; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

Qy 5 GSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQLQ 50
      || : |:| : |||||||||||||||||||||||| |
Db 7 GSPLSPEELQHQMHQHQQQQQQQQQQQQQQQQQQQQQQQQQQHQ 52

RESULT 15
T13062

CLOCK protein - fruit fly (Drosophila melanogaster)
N;Alternate names: circadian rhythm protein
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13062
R;Allada, R.; White, N.E.; So, W.V.; Hall, J.C.; Rosbash, M.
Cell 93, 791-804, 1998
A;Title: A mutant Drosophila homolog of mammalian CLOCK disrupts circadian rhythms and transcription of period and timeless.
A;Reference number: Z17596; MUID:98292177; PMID:9630223
A;Accession: T13062
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1015 <ALL>
A;Cross-references: EMBL:AF065133; NID:g3213257; PID:g3213258; PIDN:AAC39101.1
C;Genetics:
A;Gene: Clk
A;Cross-references: FlyBase:FBgn0023076
A;Map position: 3

  Query Match 46.7%; Score 134; DB 2; Length 1015;
  Best Local Similarity 100.0%; Pred. No. 1.8e-05;
  Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         24 QQQQQQQQQQQQQQQQQQQQQQQQQLQ 50
              |||||||||||||||||||||||||||
Db        798 QQQQQQQQQQQQQQQQQQQQQQQQQLQ 824
```

Search completed: March 12, 2004, 15:41:46
Job time : 10.8333 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 12, 2004, 15:39:10 ; Search time 23.4265 Seconds
(without alignments)
531.793 Million cell updates/sec

Title: US-09-620-955B-10
Perfect score: 287
Sequence: 1 LVPRGSMATLEKLMKAFESL.....QQQQQQQQQLQPGSTRAAAS 59

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description

1	208	72.5	63	14	US-10-077-584-6	Sequence 6, Appli
2	208	72.5	171	14	US-10-077-584-4	Sequence 4, Appli
3	196	68.3	91	15	US-10-354-246-1	Sequence 1, Appli
4	181	63.1	87	14	US-10-215-432-27	Sequence 27, Appl
5	181	63.1	1543	9	US-09-904-987-7	Sequence 7, Appli
6	145	50.5	338	9	US-09-933-638A-12	Sequence 12, Appl
7	145	50.5	339	15	US-10-116-275-184	Sequence 184, App
8	145	50.5	371	9	US-09-849-243-16	Sequence 16, Appl
9	144	50.2	780	9	US-09-770-689A-5	Sequence 5, Appli
10	143	49.8	1402	14	US-10-379-616-12	Sequence 12, Appl
11	142.5	49.7	548	15	US-10-437-171-4	Sequence 4, Appli
12	142.5	49.7	596	15	US-10-437-171-2	Sequence 2, Appli
13	141	49.1	966	9	US-09-801-368-372	Sequence 372, App
14	140	48.8	1572	15	US-10-116-275-179	Sequence 179, App
15	139	48.4	1138	14	US-10-074-475-194	Sequence 194, App
16	138	48.1	97	9	US-09-864-761-35499	Sequence 35499, A
17	138	48.1	1070	9	US-09-735-367B-6	Sequence 6, Appli
18	138	48.1	2005	9	US-09-735-367B-3	Sequence 3, Appli
19	138	48.1	2063	9	US-09-735-367B-2	Sequence 2, Appli
20	137	47.7	406	15	US-10-369-493-3147	Sequence 3147, Ap
21	135.5	47.2	314	14	US-10-317-832-13	Sequence 13, Appl
22	135	47.0	80	14	US-10-177-725-14	Sequence 14, Appl
23	135	47.0	910	10	US-09-086-436-31	Sequence 31, Appl
24	133	46.3	796	13	US-10-044-205A-31	Sequence 31, Appl
25	130	45.3	623	15	US-10-464-939-12	Sequence 12, Appl
26	130	45.3	1420	14	US-10-379-616-4	Sequence 4, Appli
27	130	45.3	4952	15	US-10-051-874-56	Sequence 56, Appl
28	130	45.3	5008	15	US-10-051-874-166	Sequence 166, App
29	130	45.3	5159	15	US-10-085-198-112	Sequence 112, App
30	130	45.3	5262	15	US-10-051-874-165	Sequence 165, App
31	130	45.3	5262	15	US-10-051-874-167	Sequence 167, App
32	129	44.9	907	13	US-10-008-739A-2	Sequence 2, Appli
33	129	44.9	2150	13	US-10-135-322-17	Sequence 17, Appl
34	128.5	44.8	376	15	US-10-108-260A-3233	Sequence 3233, Ap
35	128	44.6	758	9	US-09-801-368-224	Sequence 224, App
36	126.5	44.1	919	14	US-10-205-823-36	Sequence 36, Appl
37	125	43.6	467	9	US-09-416-384A-7	Sequence 7, Appli
38	123	42.9	326	14	US-10-029-386-32987	Sequence 32987, A
39	123	42.9	723	13	US-10-044-205A-32	Sequence 32, Appl
40	123	42.9	816	14	US-10-207-706-3	Sequence 3, Appli
41	121.5	42.3	398	15	US-10-374-780A-2358	Sequence 2358, Ap
42	121.5	42.3	918	15	US-10-375-592A-3	Sequence 3, Appli
43	120	41.8	59	14	US-10-177-725-8	Sequence 8, Appli
44	119	41.5	386	15	US-10-374-780A-2526	Sequence 2526, Ap
45	118	41.1	71	14	US-10-007-557-9	Sequence 9, Appli

ALIGNMENTS

RESULT 1
US-10-077-584-6
; Sequence 6, Application US/10077584
; Publication No. US20030073610A1
; GENERAL INFORMATION:
; APPLICANT: LINDQUIST, SUSAN

```
;  APPLICANT:  KROBITSCH, SYLVIA
;  APPLICANT:  OUTEIRO, TIAGO F.
;  TITLE OF INVENTION: YEAST SCREENS FOR THE TREATMENT OF HUMAN DISEASE
;  FILE REFERENCE: ARCD:367US
;  CURRENT APPLICATION NUMBER: US/10/077,584
;  CURRENT FILING DATE:  2002-02-15
;  PRIOR APPLICATION NUMBER: 60/269,157
;  PRIOR FILING DATE: 2001-02-15
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 63
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-077-584-6

  Query Match             72.5%;  Score 208;  DB 14;  Length 63;
  Best Local Similarity   95.6%;  Pred. No. 4.3e-16;
  Matches   43;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQLQP 51
              |||||||||||||||||||||||||||||||||||||||||||  |
Db          1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQPPP 45


RESULT 2
US-10-077-584-4
; Sequence 4, Application US/10077584
; Publication No. US20030073610A1
; GENERAL INFORMATION:
;  APPLICANT: LINDQUIST, SUSAN
;  APPLICANT:  KROBITSCH, SYLVIA
;  APPLICANT:  OUTEIRO, TIAGO F.
;  TITLE OF INVENTION: YEAST SCREENS FOR THE TREATMENT OF HUMAN DISEASE
;  FILE REFERENCE: ARCD:367US
;  CURRENT APPLICATION NUMBER: US/10/077,584
;  CURRENT FILING DATE:  2002-02-15
;  PRIOR APPLICATION NUMBER: 60/269,157
;  PRIOR FILING DATE: 2001-02-15
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 171
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-077-584-4

  Query Match             72.5%;  Score 208;  DB 14;  Length 171;
  Best Local Similarity   97.7%;  Pred. No. 1.3e-15;
  Matches   43;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQLQ 50
              |||||||||||||||||||||||||||||||||||||||||||| |
Db          1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 44


RESULT 3
```

```
US-10-354-246-1
; Sequence 1, Application US/10354246
; Publication No. US20030232052A1
; GENERAL INFORMATION:
;  APPLICANT: Khoshnan, Ali
;  APPLICANT:  Patterson, Paul H.
;  TITLE OF INVENTION: ANTIBODIES THAT BIND TO AN EPITOPE OF
;  TITLE OF INVENTION:  THE HUNTINGTON'S DISEASE PROTEIN
;  FILE REFERENCE: CALTE.012A
;  CURRENT APPLICATION NUMBER: US/10/354,246
;  CURRENT FILING DATE:  2003-01-28
;  PRIOR APPLICATION NUMBER: 60/353,032
;  PRIOR FILING DATE: 2001-01-28
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 91
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-354-246-1

  Query Match             68.3%;  Score 196;  DB 15;  Length 91;
  Best Local Similarity   91.1%;  Pred. No. 1.4e-14;
  Matches   41;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQP 51
              |||||||||||||||||||||||||||||||||||||||||    |
Db          1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQPPPPP 45


RESULT 4
US-10-215-432-27
; Sequence 27, Application US/10215432
; Publication No. US20030109476A1
; GENERAL INFORMATION:
;  APPLICANT: Eric B. Kmiec
;  APPLICANT:  Hetal Parekh-Olmedo
;  TITLE OF INVENTION: Composition and methods for the
;  TITLE OF INVENTION:  prevention and treatment of Huntington's disease
;  FILE REFERENCE: NaPro-10
;  CURRENT APPLICATION NUMBER: US/10/215,432
;  CURRENT FILING DATE:  2002-11-19
;  NUMBER OF SEQ ID NOS: 44
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 87
;   TYPE: PRT
;   ORGANISM: Homo Sapiens
US-10-215-432-27

  Query Match             63.1%;  Score 181;  DB 14;  Length 87;
  Best Local Similarity   69.5%;  Pred. No. 6e-13;
  Matches   41;  Conservative    0;  Mismatches    4;  Indels   14;  Gaps    1;

Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQ--------------QQQQQLQP 51
              |||||||||||||||||||||||||||||||||||              |   | ||
Db          1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQPPPPPPPPPPPQLPQPPPQAQP 59
```

```
RESULT 5
US-09-904-987-7
; Sequence 7, Application US/09904987
; Patent No. US20020037908A1
; GENERAL INFORMATION:
;  APPLICANT: No. US20020037908A1actyl, Inc.
;  TITLE OF INVENTION: Methods and Compositions for Controlling Pathological and
Prepathological
;  TITLE OF INVENTION:  Protein Assembly or Aggregation
;  FILE REFERENCE: 42108/26146
;  CURRENT APPLICATION NUMBER: US/09/904,987
;  CURRENT FILING DATE:  2001-07-12
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
;   LENGTH: 1543
;   TYPE: PRT
;   ORGANISM: homo sapiens
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: NCBI ENTREZ / XP_003405
;   DATABASE ENTRY DATE: 2001-04-16
;   RELEVANT RESIDUES: (1)..(1543)
US-09-904-987-7

  Query Match             63.1%;  Score 181;  DB 9;  Length 1543;
  Best Local Similarity   69.5%;  Pred. No. 1.3e-11;
  Matches   41;  Conservative    0;  Mismatches    4;  Indels   14;  Gaps    1;

Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQ--------------QQQQQLQP 51
              |||||||||||||||||||||||||||||||||||              |   | ||
Db          1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQPPPPPPPPPPPQLPQPPPQAQP 59


RESULT 6
US-09-933-638A-12
; Sequence 12, Application US/09933638A
; Patent No. US20020160952A1
; GENERAL INFORMATION:
;  APPLICANT: Kazantsev, Aleksey G.
;  APPLICANT:  Thompson, Leslie M.
;  APPLICANT:  Housman, David E.
;  TITLE OF INVENTION: INHIBITION OF PROTEIN-PROTEIN INTERACTION
;  FILE REFERENCE: 01997-289001
;  CURRENT APPLICATION NUMBER: US/09/933,638A
;  CURRENT FILING DATE:  2001-08-20
;  PRIOR APPLICATION NUMBER: US 60/226,502
;  PRIOR FILING DATE: 2000-08-18
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
;   LENGTH: 338
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-933-638A-12
```

```
  Query Match             50.5%;  Score 145;  DB 9;  Length 338;
  Best Local Similarity   59.3%;  Pred. No. 2.5e-08;
  Matches   32;  Conservative    7;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          6 SMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQLQPGSTRAAAS 59
              |:: ||:  :  :  :  |||||||||||||||||||||||| |     |||:
Db         48 SLSILEEQQRQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQAVAAAA 101


RESULT 7
US-10-116-275-184
; Sequence 184, Application US/10116275
; Publication No. US20030211476A1
; GENERAL INFORMATION:
;  APPLICANT: Elan Pharmaceutical Technology
;  APPLICANT:  O'Mahony, Daniel J.
;  APPLICANT:  Brayden, David
;  APPLICANT:  Byrne, Daragh
;  APPLICANT:  Lambkin, Imelda
;  APPLICANT:  Higgins, Lisa
;  TITLE OF INVENTION: Genetic Analysis of Peyer's Patches and M Cells and
Methods and
;  TITLE OF INVENTION:  Compositions Targeting Peyer's Patches and M Cell
Receptors
;  FILE REFERENCE: E1067/20087
;  CURRENT APPLICATION NUMBER: US/10/116,275
;  CURRENT FILING DATE:  2002-10-04
;  NUMBER OF SEQ ID NOS: 349
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 184
;   LENGTH: 339
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-116-275-184

  Query Match             50.5%;  Score 145;  DB 15;  Length 339;
  Best Local Similarity   59.3%;  Pred. No. 2.5e-08;
  Matches   32;  Conservative    7;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          6 SMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQLQPGSTRAAAS 59
              |:: ||:  :  :  :  |||||||||||||||||||||||| |     |||:
Db         48 SLSILEEQQRQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQAVAAAA 101


RESULT 8
US-09-849-243-16
; Sequence 16, Application US/09849243
; Patent No. US20020157127A1
;   GENERAL INFORMATION:
;        APPLICANT: Kirschbaum, Bernd
;                   Berglund, Erick
;                   Meisterernst, Michael
;                   Polites, Greg
;        TITLE OF INVENTION: PURIFICATION OF HIGHER ORDER TRANSCRIPTION
;                            COMPLEXES FROM TRANSGENIC
;                            NON-HUMAN ANIMALS
;        NUMBER OF SEQUENCES: 17
```

```
;         CORRESPONDENCE ADDRESS:
;              ADDRESSEE: HELLER, EHRMAN, WHITE & McAULIFFE
;              STREET: 1666 K Street, N.W., Suite 300
;              CITY: Washington
;              STATE: D.C.
;              COUNTRY: USA
;              ZIP: 20006
;         COMPUTER READABLE FORM:
;              MEDIUM TYPE: Floppy disk
;              COMPUTER: IBM PC compatible
;              OPERATING SYSTEM: PC-DOS/MS-DOS
;              SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/09/849,243
;              FILING DATE: 07-May-2001
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Granados, Patricia D.
;              REGISTRATION NUMBER: 33,683
;              REFERENCE/DOCKET NUMBER: 38005-0148
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: (202)912-2000
;              TELEFAX: (202)912-2020
;    INFORMATION FOR SEQ ID NO: 16:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 371 amino acids
;              TYPE: amino acid
;              TOPOLOGY: linear
;         MOLECULE TYPE: protein
;         SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-09-849-243-16

  Query Match             50.5%;  Score 145;  DB 9;  Length 371;
  Best Local Similarity   59.3%;  Pred. No. 2.7e-08;
  Matches   32;  Conservative    7;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          6 SMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQPGSTRAAAS 59
              |:: ||:  :  :  :  |||||||||||||||||||||||| |     |||:
Db         80 SLSILEEQQRQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQAVAAAA 133


RESULT 9
US-09-770-689A-5
; Sequence 5, Application US/09770689A
; Patent No. US20020115171A1
; GENERAL INFORMATION:
;  APPLICANT: YAN, Chunhua et al.
;  TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE
;  TITLE OF INVENTION:  PROTEINS, AND USES THEREOF
;  FILE REFERENCE: CL001079
;  CURRENT APPLICATION NUMBER: US/09/770,689A
;  CURRENT FILING DATE:  2001-01-29
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 780
;   TYPE: PRT
```

```
;   ORGANISM: HOMO SAPIENS
US-09-770-689A-5

  Query Match             50.2%;  Score 144;  DB 9;  Length 780;
  Best Local Similarity   80.6%;  Pred. No. 7.8e-08;
  Matches   29;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy         24 QQQQQQQQQQQQQQQQQQQQQQQQQLQPGSTRAAAS 59
              ||||||||||||||||||||||||||  ||  | ::|
Db        599 QQQQQQQQQQQQQQQQQQQQQQQQQTPGMRRCSSS 634


RESULT 10
US-10-379-616-12
; Sequence 12, Application US/10379616
; Publication No. US20030153047A1
; GENERAL INFORMATION:
;  APPLICANT: THE UNITED STATES OF AMERICA represented by THE SE
;  TITLE OF INVENTION: AIB1, A novel steriod receptor co-activator
;  FILE REFERENCE: 49944
;  CURRENT APPLICATION NUMBER: US/10/379,616
;  CURRENT FILING DATE:  2003-03-04
;  PRIOR APPLICATION NUMBER: US/09/125,635
;  PRIOR FILING DATE: 1998-08-21
;  PRIOR APPLICATION NUMBER: 60/049,728
;  PRIOR FILING DATE: 1997-06-17
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 1402
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-10-379-616-12

  Query Match             49.8%;  Score 143;  DB 14;  Length 1402;
  Best Local Similarity   62.5%;  Pred. No. 1.9e-07;
  Matches   35;  Conservative    5;  Mismatches    8;  Indels    8;  Gaps    2;

Qy          2 VPR----GSMATL----EKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQL 49
              :||    ||: ||     :|  |  ||:  |||||||||||||||||||||||:
Db        937 LPRPAMGGSVPTLPLRSNRLPGARPSLQQQQQQQQQQQQQQQQQQQQQQQQQM 992


RESULT 11
US-10-437-171-4
; Sequence 4, Application US/10437171
; Publication No. US20030235564A1
; GENERAL INFORMATION:
;  APPLICANT: Doll, Bruce
;  APPLICANT:  Fu, Huihua
;  APPLICANT:  Hollinger, Jeffrey O.
;  APPLICANT:  Sfier, Charles
;  TITLE OF INVENTION: Compositions and Devices Comprising or Encoding the Run
X2
;  TITLE OF INVENTION:  Protein and Method of Use
;  FILE REFERENCE: 1915/14014US02
;  CURRENT APPLICATION NUMBER: US/10/437,171
```

```
;  CURRENT FILING DATE:  2003-05-13
;  PRIOR APPLICATION NUMBER: 60/380,554
;  PRIOR FILING DATE: 2002-05-13
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
;   LENGTH: 548
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Peptide
US-10-437-171-4

  Query Match             49.7%;  Score 142.5;  DB 15;  Length 548;
  Best Local Similarity   58.2%;  Pred. No. 7.9e-08;
  Matches   32;  Conservative    7;  Mismatches   11;  Indels    5;  Gaps    1;

Qy          5 GSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQPGSTRAAAS 59
              | |: :  :: | :     |||||||||||||||||||||||||| |  :  |||:
Db         37 GKMSDVSPVVAAQQ-----QQQQQQQQQQQQQQQQQQQQQQQQQQEAAAAAAAA 86


RESULT 12
US-10-437-171-2
; Sequence 2, Application US/10437171
; Publication No. US20030235564A1
; GENERAL INFORMATION:
;  APPLICANT: Doll, Bruce
;  APPLICANT:  Fu, Huihua
;  APPLICANT:  Hollinger, Jeffrey O.
;  APPLICANT:  Sfier, Charles
;  TITLE OF INVENTION: Compositions and Devices Comprising or Encoding the Run
X2
;  TITLE OF INVENTION:  Protein and Method of Use
;  FILE REFERENCE: 1915/14014US02
;  CURRENT APPLICATION NUMBER: US/10/437,171
;  CURRENT FILING DATE:  2003-05-13
;  PRIOR APPLICATION NUMBER: 60/380,554
;  PRIOR FILING DATE: 2002-05-13
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
;   LENGTH: 596
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Peptide
US-10-437-171-2

  Query Match             49.7%;  Score 142.5;  DB 15;  Length 596;
  Best Local Similarity   58.2%;  Pred. No. 8.6e-08;
  Matches   32;  Conservative    7;  Mismatches   11;  Indels    5;  Gaps    1;

Qy          5 GSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQPGSTRAAAS 59
              | |: :  :: | :     |||||||||||||||||||||||||| |  :  |||:
Db        105 GKMSDVSPVVAAQQ-----QQQQQQQQQQQQQQQQQQQQQQQQQQEAAAAAAAA 154
```

RESULT 13
US-09-801-368-372
; Sequence 372, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250Alman, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 372
;   LENGTH: 966
;   TYPE: PRT
;   ORGANISM: Saccharomyces cerevisiae
US-09-801-368-372

  Query Match   49.1%; Score 141; DB 9; Length 966;
  Best Local Similarity   100.0%; Pred. No. 2.1e-07;
  Matches   28; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   24 QQQQQQQQQQQQQQQQQQQQQQQQQLQP 51
              ||||||||||||||||||||||||||||
Db   563 QQQQQQQQQQQQQQQQQQQQQQQQQLQP 590

RESULT 14
US-10-116-275-179
; Sequence 179, Application US/10116275
; Publication No. US20030211476A1
; GENERAL INFORMATION:
; APPLICANT: Elan Pharmaceutical Technology
; APPLICANT: O'Mahony, Daniel J.
; APPLICANT: Brayden, David
; APPLICANT: Byrne, Daragh
; APPLICANT: Lambkin, Imelda
; APPLICANT: Higgins, Lisa

```
;  TITLE OF INVENTION: Genetic Analysis of Peyer's Patches and M Cells and
Methods and
;  TITLE OF INVENTION:  Compositions Targeting Peyer's Patches and M Cell
Receptors
;  FILE REFERENCE: E1067/20087
;  CURRENT APPLICATION NUMBER: US/10/116,275
;  CURRENT FILING DATE:  2002-10-04
;  NUMBER OF SEQ ID NOS: 349
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 179
;   LENGTH: 1572
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-116-275-179

  Query Match             48.8%;  Score 140;  DB 15;  Length 1572;
  Best Local Similarity   67.4%;  Pred. No. 4.6e-07;
  Matches   29;  Conservative    4;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          9 TLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQLQP 51
              ||:  ::   :|   |||||||||||||||||||||| | ||
Db        201 TLQLAVQGKRTLPGLQQQQQQQQQQQQQQQQQQQQQQQPQQQP 243


RESULT 15
US-10-074-475-194
; Sequence 194, Application US/10074475
; Publication No. US20030092898A1
; GENERAL INFORMATION:
;  APPLICANT: Salceda, Susana
;  APPLICANT:  Macina, Roberto
;  APPLICANT:  Hu, Ping
;  APPLICANT:  Recipon, Herve
;  APPLICANT:  Karra, Kalpana
;  APPLICANT:  Cafferkey, Robert
;  APPLICANT:  Sun, Yongming
;  APPLICANT:  Liu, Chenghua
;  TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
;  TITLE OF INVENTION:  Genes and Proteins
;  FILE REFERENCE: DEX-0313
;  CURRENT APPLICATION NUMBER: US/10/074,475
;  CURRENT FILING DATE:  2002-02-13
;  PRIOR APPLICATION NUMBER: 60/268,292
;  PRIOR FILING DATE: 2001-02-13
;  NUMBER OF SEQ ID NOS: 295
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 194
;   LENGTH: 1138
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-10-074-475-194

  Query Match             48.4%;  Score 139;  DB 14;  Length 1138;
  Best Local Similarity   53.3%;  Pred. No. 4.2e-07;
  Matches   32;  Conservative    7;  Mismatches   17;  Indels    4;  Gaps    1;

Qy          2 VPRGSMATLEKLMKAF----ESLKSFQQQQQQQQQQQQQQQQQQQQQLQPGSTRAA 57
```

```
                ||   |:  |:| :      :  |  ||:|||||||||||||||||||||| |  | : :
Db        457 VPSSDMSPAEQLKQMAAQQQQRAKLMQQKQQQQQQQQQQQQQQQQQQQQQQQQQHSNQTS 516
```

Search completed: March 12, 2004, 15:44:13
Job time : 23.4265 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 12, 2004, 15:34:19 ; Search time 28.9216 Seconds
(without alignments)
643.657 Million cell updates/sec

Title: US-09-620-955B-10
Perfect score: 287
Sequence: 1 LVPRGSMATLEKLMKAFESL.....QQQQQQQQQLQPGSTRAAAS 59

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description

1	196	68.3	3144	4	Q9UQB7	Q9uqb7 homo sapien
2	172.5	60.1	3139	6	Q9GM99	Q9gm99 sus scrofa
3	149.5	52.1	1356	5	Q8WRE2	Q8wre2 anopheles g
4	147	51.2	556	4	O15411	O15411 homo sapien
5	147	51.2	1157	4	Q96JK7	Q96jk7 homo sapien
6	147	51.2	1761	5	O77283	O77283 drosophila
7	147	51.2	1860	5	Q8IRT3	Q8irt3 drosophila
8	147	51.2	3124	4	Q96L91	Q96191 homo sapien
9	146	50.9	2048	5	Q86JW3	Q86jw3 dictyosteli
10	145	50.5	151	4	Q7Z6S4	Q7z6s4 homo sapien
11	145	50.5	208	4	Q7Z6S5	Q7z6s5 homo sapien
12	145	50.5	653	3	Q9P788	Q9p788 schizosacch
13	145	50.5	1457	5	O44011	O44011 dictyosteli
14	144	50.2	752	11	Q8R506	Q8r506 rattus norv
15	144	50.2	780	11	Q9EQV7	Q9eqv7 rattus norv
16	144	50.2	830	11	Q99N38	Q99n38 rattus norv
17	144	50.2	858	11	Q99N37	Q99n37 rattus norv
18	143	49.8	1002	5	Q86AA4	Q86aa4 dictyosteli
19	143	49.8	1024	4	Q8IUL3	Q8iul3 homo sapien
20	143	49.8	1153	4	Q8IZL2	Q8izl2 homo sapien
21	143	49.8	1173	4	Q96JK6	Q96jk6 homo sapien
22	143	49.8	1297	5	Q8SSS5	Q8sss5 dictyosteli
23	142	49.5	1407	5	Q86H61	Q86h61 dictyosteli
24	141.5	49.3	536	3	Q9P466	Q9p466 neurospora
25	140	48.8	618	16	Q87G62	Q87g62 vibrio para
26	139.5	48.6	809	13	Q7ZVN7	Q7zvn7 brachydanio
27	139	48.4	739	11	Q7TPU6	Q7tpu6 mus musculu
28	139	48.4	1015	5	Q86AG0	Q86ag0 dictyosteli
29	139	48.4	1379	5	Q8I7P4	Q8i7p4 dictyosteli
30	139	48.4	2592	3	Q9P3J0	Q9p3j0 neurospora
31	138	48.1	398	3	Q8NJR3	Q8njr3 kluyveromyc
32	138	48.1	722	5	Q86H71	Q86h71 dictyosteli
33	138	48.1	1080	5	Q86KL1	Q86kl1 dictyosteli
34	137.5	47.9	570	11	Q9CTU8	Q9ctu8 mus musculu
35	137.5	47.9	1163	5	Q869M3	Q869m3 dictyosteli
36	137	47.7	544	4	Q9BZG7	Q9bzg7 homo sapien
37	137	47.7	1156	5	Q86HG5	Q86hg5 dictyosteli
38	137	47.7	1543	5	Q9GV71	Q9gv71 dictyosteli
39	137	47.7	1693	5	Q86JI7	Q86ji7 dictyosteli
40	137	47.7	4001	5	Q8WRQ7	Q8wrq7 drosophila
41	137	47.7	4001	5	Q9VCA8	Q9vca8 drosophila
42	136.5	47.6	149	4	Q8NFT3	Q8nft3 homo sapien
43	136.5	47.6	602	5	Q86GH6	Q86gh6 drosophila
44	136.5	47.6	1330	5	Q86GH2	Q86gh2 drosophila
45	136.5	47.6	1531	5	Q86GH1	Q86gh1 drosophila

ALIGNMENTS

```
RESULT 1
Q9UQB7
ID   Q9UQB7      PRELIMINARY;      PRT;  3144 AA.
AC   Q9UQB7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
```

```
DE   Huntingtin.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=20469406; PubMed=11013077;
RA   Matsuyama N., Hadano S., Onoe K., Osuga H., Shouguchi-Miyata J.,
RA   Gondo Y., Ikeda J.-E.;
RT   "Identification and characterization of the miniature pig Huntington's
RT   disease gene homolog: evidence for conservation and polymorphism in
RT   the CAG triplet repeat.";
RL   Genomics 69:72-85(2000).
DR   EMBL; AB016794; BAA36753.1; -.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   InterPro; IPR000091; Huntingtin.
DR   Pfam; PF03541; Huntingtin; 1.
DR   PRINTS; PR00375; HUNTINGTIN.
SQ   SEQUENCE   3144 AA;  347839 MW;  3F2BFFEFEE8E5D8E CRC64;

  Query Match             68.3%;  Score 196;  DB 4;  Length 3144;
  Best Local Similarity   91.1%;  Pred. No. 6.6e-13;
  Matches   41;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQLQP 51
              |||||||||||||||||||||||||||||||||||||||    |
Db          1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQPPPPP 45


RESULT 2
Q9GM99
ID   Q9GM99      PRELIMINARY;      PRT;  3139 AA.
AC   Q9GM99;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Huntingtin.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CSK goettingen; TISSUE=Testis;
RX   MEDLINE=20469406; PubMed=11013077;
RA   Matsuyama N., Hadano S., Onoe K., Osuga H., Shouguchi-Miyata J.,
RA   Gondo Y., Ikeda J.-E.;
RT   "Identification and characterization of the miniature pig Huntington's
RT   disease gene homolog: evidence for conservation and polymorphism in
RT   the CAG triplet repeat.";
RL   Genomics 69:72-85(2000).
DR   EMBL; AB016793; BAA36752.1; -.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   InterPro; IPR008938; ARM.
DR   InterPro; IPR001092; HLH_basic.
```

```
DR   InterPro; IPR000091; Huntingtin.
DR   Pfam; PF03541; Huntingtin; 1.
DR   PRINTS; PR00375; HUNTINGTIN.
DR   PROSITE; PS00038; HLH_1; 1.
SQ   SEQUENCE   3139 AA;  344796 MW;  051D0119A72270F8 CRC64;

  Query Match             60.1%;  Score 172.5;  DB 6;  Length 3139;
  Best Local Similarity   75.0%;  Pred. No. 2.5e-10;
  Matches   39;  Conservative    0;  Mismatches    6;  Indels    7;  Gaps    1;

Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQ-------QQQQQQQLQP 51
              |||||||||||||||||||||||||||||||||       |  |   | ||
Db          1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQLPPPPPQPPQPPPQTQP 52


RESULT 3
Q8WRE2
ID   Q8WRE2      PRELIMINARY;      PRT;  1356 AA.
AC   Q8WRE2;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Trex.
OS   Anopheles gambiae (African malaria mosquito).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX   NCBI_TaxID=7165;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Luna C., Wang X., Huang Y., Zhang J., Zheng L.;
RT   "Characterization of Four Toll Related Genes During Development and
RT   Immune Responses in Anopheles gambiae.";
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF444783; AAL37904.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0004909; F:interleukin-1, Type I, activating receptor . . .; IEA.
DR   GO; GO:0004888; F:transmembrane receptor activity; IEA.
DR   InterPro; IPR004075; IL1_receptor1.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003591; LRR_typ.
DR   InterPro; IPR002155; Thiolase.
DR   InterPro; IPR000157; TIR.
DR   InterPro; IPR008197; WAP.
DR   Pfam; PF00560; LRR; 19.
DR   Pfam; PF01582; TIR; 1.
DR   PRINTS; PR01537; INTRLKN1R1F.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 3.
DR   SMART; SM00255; TIR; 1.
DR   PROSITE; PS00317; 4_DISULFIDE_CORE; 1.
DR   PROSITE; PS00098; THIOLASE_1; 1.
DR   PROSITE; PS50104; TIR; 1.
SQ   SEQUENCE   1356 AA;  154545 MW;  DEF801EC2302ECDA CRC64;

  Query Match             52.1%;  Score 149.5;  DB 5;  Length 1356;
```

```
  Best Local Similarity   55.1%;  Pred. No. 3.9e-08;
  Matches   38;  Conservative    5;  Mismatches   11;  Indels   15;  Gaps    2;

Qy          3 PRGSM-------ATLEKL--------MKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQ 47
              |||         || : |        |:  :  :| ||||||||||||||||||||||||
Db       1265 PRGPQGRSTDYHATQQPLPLPGLASEMQPQQLHRSQQQQQQQQQQQQQQQQQQQQQQQQ 1324

Qy         48 QLQPGSTRA 56
              | || ||:|
Db       1325 QHQPPSTQA 1333


RESULT 4
O15411
ID   O15411      PRELIMINARY;      PRT;   556 AA.
AC   O15411;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CAGH32 (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97369492; PubMed=9225980;
RA   Margolis R.L., Abraham M.R., Gatchell S.B., Li S.H., Kidwai A.S.,
RA   Breschel T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;
RT   "cDNAs with long CAG trinucleotide repeats from human brain.";
RL   Hum. Genet. 100:114-122(1997).
DR   EMBL; U80743; AAB91441.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   556 AA;  57588 MW;  AAAF9DFEF777EE9E CRC64;

  Query Match             51.2%;  Score 147;  DB 4;  Length 556;
  Best Local Similarity   60.0%;  Pred. No. 3.2e-08;
  Matches   36;  Conservative    5;  Mismatches   13;  Indels    6;  Gaps    1;

Qy          1 LVPRGSMATLEKL------MKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQLQPGST 54
              |||: | ||  :|        |: |:  |||||||||||||||||||||||| |  :|
Db        267 LVPQVSQATGVQLPGKTITPAHFQLLRQQQQQQQQQQQQQQQQQQQQQQQQQQQTTT 326


RESULT 5
Q96JK7
ID   Q96JK7      PRELIMINARY;      PRT;  1157 AA.
AC   Q96JK7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein KIAA1818 (Fragment).
GN   KIAA1818.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=21245130; PubMed=11347906;
RA   Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XX.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large Proteins in vitro.";
RL   DNA Res. 8:85-95(2001).
DR   EMBL; AB058721; BAB47447.1; -.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   InterPro; IPR001005; Myb_DNA_binding.
DR   SMART; SM00717; SANT; 1.
DR   PROSITE; PS50090; MYB_3; 1.
KW   Hypothetical protein.
FT   NON_TER       1       1
SQ   SEQUENCE   1157 AA;  125525 MW;  B08A6AE50B1A9E01 CRC64;

  Query Match             51.2%;  Score 147;  DB 4;  Length 1157;
  Best Local Similarity   60.0%;  Pred. No. 6.4e-08;
  Matches   36;  Conservative    5;  Mismatches   13;  Indels    6;  Gaps    1;

Qy          1 LVPRGSMATLEKL------MKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQPGST 54
              |||: | ||  :|         |: |:  |||||||||||||||||||||||| |  :|
Db        726 LVPQVSQATGVQLPGKTITPAHFQLLRQQQQQQQQQQQQQQQQQQQQQQQQQQQTTT 785


RESULT 6
O77283
ID   O77283      PRELIMINARY;      PRT;  1761 AA.
AC   O77283;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   EG:EG0002.3 protein.
GN   EG:EG0002.3 OR CG2904.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
```

```
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Bolshakov V., Borkova D., Minana B., Kafatos F.;
RT   "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Benos P.;
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003429; AAF45898.1; -.
DR   EMBL; AL031130; CAA20016.1; -.
DR   PIR; T13675; T13675.
DR   FlyBase; FBgn0025376; EG:EG0002.3.
DR   GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR   GO; GO:0004221; F:ubiquitin thiolesterase activity; IEA.
DR   GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
DR   InterPro; IPR001394; Peptidase_C19.
DR   Pfam; PF00443; UCH; 1.
SQ   SEQUENCE   1761 AA;  192843 MW;  BB300CC95D38EB77 CRC64;

  Query Match             51.2%;  Score 147;  DB 5;  Length 1761;
  Best Local Similarity   60.4%;  Pred. No. 9.5e-08;
  Matches   29;  Conservative    8;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          3 PRGSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQLQ 50
              | |: | ::: ::  :  :  |||||||||||||||||||||||| |
Db       1474 PAGATADMQRYVQRMQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 1521
```

```
RESULT 7
Q8IRT3
ID   Q8IRT3      PRELIMINARY;      PRT;  1860 AA.
AC   Q8IRT3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG2904-PB.
GN   EG:EG0002.3 OR CG2904.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
```

```
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003429; AAN09109.1; -.
DR   FlyBase; FBgn0025376; EG:EG0002.3.
DR   GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR   GO; GO:0004221; F:ubiquitin thiolesterase activity; IEA.
DR   GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
DR   InterPro; IPR001394; Peptidase_C19.
DR   Pfam; PF00443; UCH; 1.
SQ   SEQUENCE   1860 AA;  203948 MW;  84ABE9216C6AC6E5 CRC64;

  Query Match             51.2%;  Score 147;  DB 5;  Length 1860;
  Best Local Similarity   60.4%;  Pred. No. 1e-07;
  Matches   29;  Conservative    8;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          3 PRGSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQ 50
              | |: | ::: ::  :  :  |||||||||||||||||||||||||| |
Db       1573 PAGATADMQRYVQRMQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 1620


RESULT 8
Q96L91
ID   Q96L91      PRELIMINARY;      PRT;  3124 AA.
AC   Q96L91;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
```

```
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   P400 SWI2/SNF2-related protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21400441; PubMed=11509179;
RA   Fuchs M., Gerber J., Drapkin R., Sif S., Ikura T., Ogryzko V.,
RA   Lane W.S., Nakatani Y., Livingston D.M.;
RT   "The p400 complex is an essential E1A transformation target.";
RL   Cell 106:297-307(2001).
DR   EMBL; AY044869; AAK97789.1; -.
DR   Genew; HGNC:11958; EP400.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR006562; HSA.
DR   InterPro; IPR001005; Myb_DNA_binding.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 2.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00573; HSA; 1.
DR   SMART; SM00717; SANT; 1.
DR   PROSITE; PS50090; MYB_3; 1.
KW   ATP-binding; Helicase; Hydrolase.
SQ   SEQUENCE   3124 AA;  340146 MW;  E8F57FD6C7BD01E9 CRC64;

  Query Match             51.2%;  Score 147;  DB 4;  Length 3124;
  Best Local Similarity   60.0%;  Pred. No. 1.6e-07;
  Matches   36;  Conservative    5;  Mismatches   13;  Indels    6;  Gaps    1;

Qy          1 LVPRGSMATLEKL------MKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQLQPGST 54
              |||: | ||  :|         |: |:  |||||||||||||||||||||||| |  :|
Db       2693 LVPQVSQATGVQLPGKTITPAHFQLLRQQQQQQQQQQQQQQQQQQQQQQQQQQQTTT 2752


RESULT 9
Q86JW3
ID   Q86JW3      PRELIMINARY;      PRT;  2048 AA.
AC   Q86JW3;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=AX4;
RX   MEDLINE=22092622; PubMed=12097910;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and analysis of chromosome 2 of Dictyostelium discoideum.";
RL   Nature 418:79-85(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Baumgart C.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC116984; AAO51396.1; -.
DR   InterPro; IPR008938; ARM.
DR   InterPro; IPR000904; Sec7.
DR   Pfam; PF01369; Sec7; 1.
DR   SMART; SM00222; Sec7; 1.
DR   PROSITE; PS50190; SEC7; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   2048 AA;  231362 MW;  7F7F34A35CAB8DB2 CRC64;

  Query Match             50.9%;  Score 146;  DB 5;  Length 2048;
  Best Local Similarity   62.5%;  Pred. No. 1.4e-07;
  Matches   30;  Conservative    7;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          6 SMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQPGS 53
              |:: ||:|  ::  ::  ||   |||||||||||||||||||||| |
Db        988 SISFLERLRVSYLGVEQQQQSNSQQQQQQQQQQQQQQQQQQQQQLQPNS 1035


RESULT 10
Q7Z6S4
ID   Q7Z6S4      PRELIMINARY;      PRT;   151 AA.
AC   Q7Z6S4;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   DJ191N21.2.3 (TATA box binding protein (GTF2D, SCA17, TFIID), variant
DE   3) (Fragment).
GN   TBP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Griffiths C.;
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL031259; CAD92544.1; -.
KW   Proteasome.
FT   NON_TER     151    151
SQ   SEQUENCE   151 AA;  16659 MW;  F53926CE2BAC5E6C CRC64;

  Query Match             50.5%;  Score 145;  DB 4;  Length 151;
  Best Local Similarity   59.3%;  Pred. No. 1.6e-08;
  Matches   32;  Conservative    7;  Mismatches   15;  Indels    0;  Gaps    0;
```

```
Qy          6 SMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQLQPGSTRAAAS 59
              |:: ||:  :  :  :  |||||||||||||||||||||||| |     |||:
Db         48 SLSILEEQQRQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQAVAAAA 101


RESULT 11
Q7Z6S5
ID   Q7Z6S5      PRELIMINARY;      PRT;   208 AA.
AC   Q7Z6S5;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   DJ191N21.2.2 (TATA box binding protein (GTF2D, SCA17, TFIID), variant
DE   2) (Fragment).
GN   TBP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Griffiths C.;
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL031259; CAD92543.1; -.
KW   Proteasome.
FT   NON_TER     208    208
SQ   SEQUENCE   208 AA;  22921 MW;  95792234163A9618 CRC64;

  Query Match             50.5%;  Score 145;  DB 4;  Length 208;
  Best Local Similarity   59.3%;  Pred. No. 2.1e-08;
  Matches   32;  Conservative    7;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          6 SMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQLQPGSTRAAAS 59
              |:: ||:  :  :  :  |||||||||||||||||||||||| |     |||:
Db         48 SLSILEEQQRQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQAVAAAA 101


RESULT 12
Q9P788
ID   Q9P788      PRELIMINARY;      PRT;   653 AA.
AC   Q9P788;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Putative transcriptional regulatory protein (Fragment).
GN   SPBP35G2.15.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972h-;
RA   Seeger K., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL163702; CAB87377.1; -.
```

```
FT   NON_TER      653     653
SQ   SEQUENCE   653 AA;  73291 MW;  C88A9EEDECF8B80F CRC64;

  Query Match             50.5%;  Score 145;  DB 3;  Length 653;
  Best Local Similarity   62.5%;  Pred. No. 6.2e-08;
  Matches   30;  Conservative    6;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          3 PRGSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQ 50
              |   ::  :  :: | ||:  |||||||||||||||||||||||| |
Db        236 PARLISIYQNQIQKFRSLQHMQQQQQQQQQQQQQQQQQQQQQQQQQQQ 283


RESULT 13
O44011
ID   O44011      PRELIMINARY;      PRT;  1457 AA.
AC   O44011;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Protein kinase YakA.
GN   YAKA.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AK800;
RX   MEDLINE=96042901; PubMed=8536963;
RA   Loomis W.F., Welker D., Hughes J., Maghakian D., Kuspa A.;
RT   "Integrated maps of the chromosomes in Dictyostelium discoideum.";
RL   Genetics 141:147-157(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AK800;
RX   MEDLINE=96224325; PubMed=8643615;
RA   Kuspa A., Loomis W.F.;
RT   "Ordered yeast artificial chromosome clones representing the
RT   Dictyostelium discoideum genome.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:5562-5566(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AK800;
RA   Kuspa A., Lu S., Souza G.M.;
RT   "YakA, a protein kinase required for the growth to development
RT   transition in Dictyostelium.";
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR   EMBL; AF045453; AAC02554.1; -.
DR   PIR; T14577; T14577.
DR   HSSP; P24941; 1CKP.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
```

```
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ   SEQUENCE   1457 AA;  167111 MW;  C1FCDCE99D561856 CRC64;

  Query Match             50.5%;  Score 145;  DB 5;  Length 1457;
  Best Local Similarity   59.2%;  Pred. No. 1.3e-07;
  Matches   29;  Conservative    7;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          2 VPRGSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQLQ 50
              :|: ||    :::   :  :  |||||||||||||||||||||||| |
Db        575 IPQHSMLNGNQILNQHQLFQQLQQQQQQQQQQQQQQQQQQQQQQQQQQ 623


RESULT 14
Q8R506
ID   Q8R506      PRELIMINARY;      PRT;   752 AA.
AC   Q8R506;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Nadrin-102.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Harada A., Furuta B., Takeuchi K., Itakura M., Takahashi M., Umeda M.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20538431; PubMed=10967100;
RA   Harada A., Furuta B., Takeuchi K., Itakura M., Takahashi M., Umeda M.;
RT   "Nadrin, a Novel Neuron-specific GTPase-activating Protein Involved in
RT   Regulated Exocytosis.";
RL   J. Biol. Chem. 275:36885-36891(2000).
DR   EMBL; AB080637; BAB85655.1; -.
DR   InterPro; IPR006632; BAR.
DR   InterPro; IPR000198; RhoGAP.
DR   InterPro; IPR008936; Rho_GAP.
DR   Pfam; PF00620; RhoGAP; 1.
DR   SMART; SM00721; BAR; 1.
DR   SMART; SM00324; RhoGAP; 1.
DR   PROSITE; PS50238; RHOGAP; 1.
SQ   SEQUENCE   752 AA;  82520 MW;  D9002F74E5BD1AE1 CRC64;

  Query Match             50.2%;  Score 144;  DB 11;  Length 752;
  Best Local Similarity   80.6%;  Pred. No. 9.1e-08;
  Matches   29;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy         24 QQQQQQQQQQQQQQQQQQQQQQQQQLQPGSTRAAAS 59
              |||||||||||||||||||||||||  ||  | ::|
```

```
Db        599 QQQQQQQQQQQQQQQQQQQQQQQQQQTPGMRRCSSS 634


RESULT 15
Q9EQV7
ID   Q9EQV7      PRELIMINARY;      PRT;   780 AA.
AC   Q9EQV7;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Nadrin.
GN   NADRIN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=20538431; PubMed=10967100;
RA   Harada A., Furuta B., Takeuchi K., Itakura M., Takahashi M., Umeda M.;
RT   "Nadrin, a Novel Neuron-specific GTPase-activating Protein Involved in
RT   Regulated Exocytosis.";
RL   J. Biol. Chem. 275:36885-36891(2000).
DR   EMBL; AB042827; BAB12426.1; -.
DR   HSSP; Q07960; 1RGP.
DR   InterPro; IPR006632; BAR.
DR   InterPro; IPR000198; RhoGAP.
DR   InterPro; IPR008936; Rho_GAP.
DR   Pfam; PF00620; RhoGAP; 1.
DR   SMART; SM00721; BAR; 1.
DR   SMART; SM00324; RhoGAP; 1.
DR   PROSITE; PS50238; RHOGAP; 1.
SQ   SEQUENCE   780 AA;  85824 MW;  180B75771C510246 CRC64;

  Query Match             50.2%;  Score 144;  DB 11;  Length 780;
  Best Local Similarity   80.6%;  Pred. No. 9.5e-08;
  Matches   29;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy         24 QQQQQQQQQQQQQQQQQQQQQQQQQLQPGSTRAAAS 59
              |||||||||||||||||||||||||  ||  | ::|
Db        599 QQQQQQQQQQQQQQQQQQQQQQQQQQTPGMRRCSSS 634


Search completed: March 12, 2004, 15:41:00
Job time : 30.9216 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 12, 2004, 15:22:04 ; Search time 6.94118 Seconds
(without alignments)
442.596 Million cell updates/sec

Title: US-09-620-955B-10
Perfect score: 287
Sequence: 1 LVPRGSMATLEKLMKAFESL.....QQQQQQQQQLQPGSTRAAAS 59

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	196	68.3	3144	1	HD_HUMAN	P42858 homo sapien
2	147	51.2	1081	1	GALY_YEAST	P19659 saccharomyc
3	145	50.5	339	1	TBP_HUMAN	P20226 homo sapien
4	142.5	49.7	607	1	RUN2_MOUSE	Q08775 m runt-rela
5	141	49.1	966	1	SSN6_YEAST	P14922 saccharomyc
6	140.5	49.0	590	1	HMAA_DROME	P29555 drosophila
7	138	48.1	2063	1	NCO6_HUMAN	Q14686 h nuclear r
8	137	47.7	376	1	MJD1_HUMAN	P54252 homo sapien
9	136.5	47.6	2067	1	NCO6_MOUSE	Q9jl19 m nuclear r
10	136	47.4	313	1	THAB_HUMAN	Q96ek4 homo sapien
11	136	47.4	714	1	FXP2_MOUSE	P58463 mus musculu
12	136	47.4	715	1	FXP2_HUMAN	O15409 homo sapien
13	136	47.4	716	1	FXP2_PANTR	Q8mja0 pan troglod
14	136	47.4	1319	1	MN1_HUMAN	Q10571 homo sapien
15	135	47.0	910	1	HCN1_MOUSE	O88704 mus musculu
16	135	47.0	1167	1	WC1_NEUCR	Q01371 neurospora
17	135	47.0	1177	1	SP97_DICDI	Q95zg3 dictyosteli

18	135	47.0	1516	1	NCO2_XENLA	Q9w705 xenopus lae
19	134	46.7	1023	1	CLOC_DROME	O61735 drosophila
20	133.5	46.5	445	1	PO32_MOUSE	P31360 mus musculu
21	133.5	46.5	445	1	PO32_RAT	P56222 rattus norv
22	133	46.3	796	1	CN04_HUMAN	Q9h1b7 homo sapien
23	133	46.3	905	1	SNF5_YEAST	P18480 saccharomyc
24	132.5	46.2	1586	1	SN22_HUMAN	P51531 homo sapien
25	132	46.0	1398	1	NCO3_MOUSE	O09000 m nuclear r
26	131	45.6	705	1	FXP1_MOUSE	P58462 mus musculu
27	131	45.6	2212	1	T230_HUMAN	Q93074 homo sapien
28	130	45.3	623	1	DSH_DROME	P51140 drosophila
29	130	45.3	1424	1	NCO3_HUMAN	Q9y6q9 h nuclear r
30	130	45.3	5262	1	MLL2_HUMAN	O14686 homo sapien
31	129	44.9	521	1	RUN2_HUMAN	Q13950 h runt-rela
32	129	44.9	907	1	ANDR_CANFA	Q9tt90 canis famil
33	129	44.9	1090	1	NIT4_NEUCR	P28349 neurospora
34	129	44.9	1905	1	TAGB_DICDI	P54683 dictyosteli
35	129	44.9	2703	1	NOTC_DROME	P07207 drosophila
36	128	44.6	443	1	PO32_HUMAN	P20265 homo sapien
37	128	44.6	758	1	YM38_YEAST	Q03825 saccharomyc
38	128	44.6	910	1	HCN1_RAT	Q9jkb0 rattus norv
39	128	44.6	1161	1	BM2K_HUMAN	Q9nsy1 homo sapien
40	127	44.3	644	1	BTD_DROME	Q24266 drosophila
41	127	44.3	700	1	BIB_DROME	P23645 drosophila
42	126.5	44.1	919	1	ANDR_HUMAN	P10275 homo sapien
43	125	43.6	902	1	ANDR_RAT	P15207 rattus norv
44	125	43.6	1012	1	PHC1_MOUSE	Q64028 mus musculu
45	125	43.6	3726	1	ABF1_MOUSE	Q61329 mus musculu

ALIGNMENTS

```
RESULT 1
HD_HUMAN
ID   HD_HUMAN        STANDARD;      PRT;  3144 AA.
AC   P42858;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Huntingtin (Huntington's disease protein) (HD protein).
GN   HD OR IT15.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Retina;
RX   MEDLINE=93208892; PubMed=8458085;
RA   Macdonald M., Ambrose C.M., Duyao M.P., Myers R.H., Lin C.S.,
RA   Srinidhi J., Barnes G., Taylor S.A., James M., Groot N., McFarlane H.,
RA   Jenkins B., Anderson M.A., Wexler N.S., Gusella J.F., Bates G.P.,
RA   Baxendale S., Hummerich H., Kirby S., North M., Youngman S., Mott R.,
RA   Zehetner G., Sedlacek Z., Poustka A., Frischauf A.-M., Lehrach H.,
RA   Buckler A.J., Church D., Doucette-Stamm L., O'Donovan M.C.,
RA   Riba-Ramirez L., Shah M., Stanton V.P., Strobel S.A., Draths K.M.,
```

RA Wales J.L., Dervan P., Housman D.E., Altherr M., Shiang R.,
RA Thompson L., Fielder T., Wasmuth J.J., Tagle D., Valdes J., Elmer L.,
RA Allard M., Castilla L., Swaroop M., Blanchard K., Collins F.S.,
RA Snell R., Holloway T., Gillespie K., Datson N., Shaw S., Harper P.S.;
RT "A novel gene containing a trinucleotide repeat that is expanded and
RT unstable on Huntington's disease chromosomes. The Huntington's
RT Disease Collaborative Research Group.";
RL Cell 72:971-983(1993).
RN [2]
RP SEQUENCE OF 1-90 FROM N.A.
RX MEDLINE=95278941; PubMed=7759106;
RA Lin B., Nasir J., Kalchman M.A., McDonald H., Zeisler J.,
RA Goldberg Y.P., Hayden M.R.;
RT "Structural analysis of the 5' region of mouse and human Huntington
RT disease genes reveals conservation of putative promoter region and
RT di- and trinucleotide polymorphisms.";
RL Genomics 25:707-715(1995).
RN [3]
RP SEQUENCE OF 1-205 FROM N.A.
RX MEDLINE=94255787; PubMed=8197474;
RA Ambrose C.M., Duyao M.P., Barnes G., Bates G.P., Lin C.S.,
RA Srinidhi J., Baxendale S., Hummerich H., Lehrach H., Altherr M.,
RA Wasmuth J., Buckler A., Church D., Housman D., Berks M., Micklem G.,
RA Durbin R., Dodge A., Read A., Gusella J.F., Macdonald M.E.;
RT "Structure and expression of the Huntington's disease gene: evidence
RT against simple inactivation due to an expanded CAG repeat.";
RL Somat. Cell Mol. Genet. 20:27-38(1994).
RN [4]
RP SEQUENCE OF 1-117 FROM N.A.
RA Matthews P.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN [5]
RP SEQUENCE OF 119-934 FROM N.A.
RA Lloyd C.;
RL Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
RN [6]
RP SEQUENCE OF 1212-1290 FROM N.A.
RA Mungall A., Odell C.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
RN [7]
RP SEQUENCE OF 1291-1860 FROM N.A.
RA Mungall A.;
RL Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
RN [8]
RP SEQUENCE OF 1862-2820 FROM N.A.
RA Buck D.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
RN [9]
RP SEQUENCE OF 2563-3144 FROM N.A.
RC TISSUE=Brain, Caudate, Frontal cortex, Muscle, and Retina;
RX MEDLINE=94093536; PubMed=7903579;
RA Lin B., Rommens J.M., Graham R.K., Kalchman M., Macdonald H.,
RA Nasir J., Delaney A., Goldberg Y.P., Hayden M.R.;
RT "Differential 3' polyadenylation of the Huntington disease gene
RT results in two mRNA species with variable tissue expression.";
RL Hum. Mol. Genet. 2:1541-1545(1993).
RN [10]

```
RP   SUBCELLULAR LOCATION.
RX   MEDLINE=95375771; PubMed=7647777;
RA   Trottier Y., Devys D., Imbert G., Saudou F., An I., Lutz Y., Weber C.,
RA   Agid Y., Hirsch E.C., Mandel J.-L.;
RT   "Cellular localization of the Huntington's disease protein and
RT   discrimination of the normal and mutated form.";
RL   Nat. Genet. 10:104-110(1995).
RN   [11]
RP   CLEAVAGE BY APOPAIN.
RX   MEDLINE=96331285; PubMed=8696339;
RA   Goldberg Y.P., Nicholson D.W., Rasper D.M., Kalchman M.A., Koide H.B.,
RA   Graham R.K., Bromm M., Kazemi-Esfarjani P., Thornberry N.A.,
RA   Vaillancourt J.P., Hayden M.R.;
RT   "Cleavage of huntingtin by apopain, a proapoptotic cysteine protease,
RT   is modulated by the polyglutamine tract.";
RL   Nat. Genet. 13:442-449(1996).
RN   [12]
RP   INTERACTION WITH FNBP3.
RX   MEDLINE=98367036; PubMed=9700202;
RA   Faber P.W., Barnes G.T., Srinidhi J., Chen J., Gusella J.F.,
RA   MacDonald M.E.;
RT   "Huntingtin interacts with a family of WW domain proteins.";
RL   Hum. Mol. Genet. 7:1463-1474(1998).
CC   -!- FUNCTION: May play a role in microtubule-mediated transport or
CC       vesicle function.
CC   -!- SUBUNIT: Binds SH3GLB1 (By similarity). Interacts through its N-
CC       terminus with FNBP3.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- TISSUE SPECIFICITY: Widely expressed with the highest level of
CC       expression in the brain (nerve fibers, varicosities, and nerve
CC       endings). In the brain, the regions where it can be mainly found
CC       are the cerebellar cortex, the neocortex, the striatum, and the
CC       hippocampal formation.
CC   -!- PTM: Cleaved by apopain downstream of the polyglutamine stretch.
CC       The resulting amino-terminal fragment is cytotoxic and provokes
CC       apoptosis.
CC   -!- POLYMORPHISM: The poly-Gln region of HD is highly polymorphic (10
CC       to 35 repeats) in the normal population and is expanded to about
CC       36-120 repeats in hd patients. The repeat length usually increases
CC       in successive generations, but contracts also on occasion. The
CC       longer expansions result in earlier onset and more severe clinical
CC       manifestations of the disease. The adjacent poly-pro region is
CC       also polymorphic and varies between 7-12 residues. Polyglutamine
CC       expansion leads to elevated susceptibility to apopain cleavage and
CC       likely result in accelerated neuronal apoptosis.
CC   -!- DISEASE: DEFECTS IN HD ARE THE CAUSE OF HUNTINGTON'S DISEASE, AN
CC       AUTOSOMAL DOMINANT NEURODEGENERATIVE DISORDER CHARACTERIZED BY
CC       INVOLUNTARY MOVEMENTS (CHOREA), GENERAL MOTOR IMPAIRMENT,
CC       PSYCHIATRIC DISORDERS AND DEMENTIA. ONSET OF THE DISEASE OCCURS
CC       USUALLY IN THE THIRD OR FOURTH DECADE OF LIFE AND SYMPTOMS
CC       PROGRESSIVELY WORSEN LEADING TO DEATH IN 10 TO 20 YEARS. IT
CC       AFFECTS 1 IN 10,000 INDIVIDUALS OF EUROPEAN ORIGIN. NEUROPATHOLOGY
CC       OF HUNTINGTON'S DISEASE DISPLAYS A DISTINCTIVE PATTERN WITH LOSS
CC       OF NEURONS, SPECIALLY IN THE CAUDATE AND PUTAMEN (STRIATUM).
CC   -!- SIMILARITY: Contains 10 HEAT repeats.
CC   -!- SIMILARITY: Belongs to the hungtintin family.
CC   -!- DATABASE: NAME=HotMolecBase; NOTE=HD entry;
```

```
CC
WWW="http://bioinformatics.weizmann.ac.il/hotmolecbase/entries/hunti.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L12392; AAB38240.1; -.
DR   EMBL; L34020; -; NOT_ANNOTATED_CDS.
DR   EMBL; L27350; -; NOT_ANNOTATED_CDS.
DR   EMBL; L27351; -; NOT_ANNOTATED_CDS.
DR   EMBL; L27352; -; NOT_ANNOTATED_CDS.
DR   EMBL; L27353; -; NOT_ANNOTATED_CDS.
DR   EMBL; L27354; -; NOT_ANNOTATED_CDS.
DR   EMBL; Z68756; -; NOT_ANNOTATED_CDS.
DR   EMBL; Z49155; CAA89025.1; -.
DR   EMBL; Z49208; -; NOT_ANNOTATED_CDS.
DR   EMBL; Z69649; -; NOT_ANNOTATED_CDS.
DR   EMBL; Z49154; CAA89024.1; -.
DR   EMBL; Z49769; CAA89839.1; -.
DR   EMBL; L20431; AAA52702.1; -.
DR   PIR; A46068; A46068.
DR   Genew; HGNC:4851; HD.
DR   MIM; 143100; -.
DR   GO; GO:0005737; C:cytoplasm; TAS.
DR   GO; GO:0005634; C:nucleus; TAS.
DR   GO; GO:0005625; C:soluble fraction; TAS.
DR   GO; GO:0008017; F:microtubule binding; TAS.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   GO; GO:0003714; F:transcription co-repressor activity; TAS.
DR   GO; GO:0005215; F:transporter activity; TAS.
DR   GO; GO:0007610; P:behavior; TAS.
DR   GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR   GO; GO:0006917; P:induction of apoptosis; TAS.
DR   GO; GO:0009405; P:pathogenesis; TAS.
DR   InterPro; IPR000091; Huntingtin.
DR   Pfam; PF03541; Huntingtin; 1.
DR   PRINTS; PR00375; HUNTINGTIN.
KW   Repeat; Disease mutation; Polymorphism; Triplet repeat expansion;
KW   Apoptosis.
FT   DOMAIN      205    329       HEAT REPEATS DOMAIN 1.
FT   DOMAIN      745    942       HEAT REPEATS DOMAIN 2.
FT   DOMAIN     1534   1575       HEAT REPEATS DOMAIN 3.
FT   DOMAIN       18     40       POLY-GLN.
FT   DOMAIN       41     51       POLY-PRO.
FT   DOMAIN       65     80       POLY-PRO.
FT   DOMAIN     1439   1442       POLY-THR.
FT   DOMAIN     2343   2347       POLY-GLU.
FT   DOMAIN     2640   2645       POLY-GLU.
FT   SITE        513    514       CLEAVAGE (BY APOPAIN) (POTENTIAL).
FT   SITE        530    531       CLEAVAGE (BY APOPAIN) (POTENTIAL).
FT   SITE        552    553       CLEAVAGE (BY APOPAIN) (POTENTIAL).
FT   SITE        589    590       CLEAVAGE (BY APOPAIN) (POTENTIAL).
```

```
FT   VARIANT      38     40       Missing.
FT                                /FTId=VAR_005268.
FT   CONFLICT   2788   2788       V -> I (IN REF. 10).
SQ   SEQUENCE   3144 AA;  347855 MW;  9D1BA8528929908F CRC64;

  Query Match             68.3%;  Score 196;  DB 1;  Length 3144;
  Best Local Similarity   91.1%;  Pred. No. 1.6e-10;
  Matches   41;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQP 51
              |||||||||||||||||||||||||||||||||||||||||    |
Db          1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQPPPPP 45


RESULT 2
GALY_YEAST
ID   GALY_YEAST     STANDARD;      PRT;  1081 AA.
AC   P19659; Q08221;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Transcription regulatory protein GAL11.
GN   GAL11 OR SPT13 OR RAR3 OR YOL051W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=89096873; PubMed=3062377;
RA   Suzuki Y., Nogi Y., Abe A., Fukasawa T.;
RT   "GAL11 protein, an auxiliary transcription activator for genes
RT   encoding galactose-metabolizing enzymes in Saccharomyces
RT   cerevisiae.";
RL   Mol. Cell. Biol. 8:4991-4999(1988).
RN   [2]
RP   REVISIONS.
RX   MEDLINE=93024425; PubMed=1406662;
RA   Suzuki Y., Nogi Y., Abe A., Fukasawa T.;
RL   Mol. Cell. Biol. 12:4806-4806(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Ansorge W., Benes V., Rechmann S., Schwager C., Teodoru C.,
RA   Voss H., Wiemann S.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 1-352 FROM N.A.
RC   STRAIN=S288c / FY73;
RX   MEDLINE=96381248; PubMed=8789261;
RA   Mannhaupt G., Vetter I., Schwarzlose C., Mitzel S., Feldmann H.;
RT   "Analysis of a 26 kb region on the left arm of yeast chromosome XV.";
RL   Yeast 12:67-76(1996).
RN   [5]
RP   CHARACTERIZATION.
RX   MEDLINE=91172223; PubMed=2005915;
RA   Long R.M., Mylin L.M., Hopper J.E.;
```

```
RT   "GAL11 (SPT13), a transcriptional regulator of diverse yeast genes,
RT   affects the phosphorylation state of GAL4, a highly specific
RT   transcriptional activator.";
RL   Mol. Cell. Biol. 11:2311-2314(1991).
CC   -!- FUNCTION: Auxiliary transcription activator for genes encoding
CC       galactose-metabolizing enzymes. Essential for normal growth on
CC       nonfermentable carbon sources, for sporulation and mating.
CC       Coactivator that links transcriptional activators such as GAL4
CC       and GRF1/RAP1/TUF1 with the basic transcription machinery,
CC       possibly by protein-protein interactions.
CC   -!- FUNCTION: It has an important role in the negative regulation of
CC       Ty transcription.
CC   -!- MISCELLANEOUS: GAL11 lacks a DNA-domain, it probably complexes
CC       with GAL4 that has the capacity to bind DNA. Association between
CC       GAL11 and GAL4 may serve to expedite phosphorylation of GAL4.
CC   -!- SIMILARITY: TO K.LACTIS GALY, AND SOME, TO YEAST GLUCOSE
CC       REPRESSION MEDIATOR PROTEIN (CYC8).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M22481; AAA34622.1; -.
DR   EMBL; Z74793; CAA99056.1; -.
DR   EMBL; X91067; CAA62537.1; -.
DR   PIR; S66736; S66736.
DR   GermOnline; 143473; -.
DR   TRANSFAC; T03313; -.
DR   SGD; S0005411; GAL11.
DR   GO; GO:0000119; C:mediator complex; IDA.
DR   GO; GO:0016455; F:RNA polymerase II transcription mediator ac. . .; IDA.
DR   GO; GO:0006366; P:transcription from Pol II promoter; IDA.
DR   InterPro; IPR008626; GAL11.
DR   Pfam; PF05397; GAL11; 1.
KW   Transcription regulation; Activator; Galactose metabolism;
KW   Repeat.
FT   DOMAIN      147    158       POLY-GLN.
FT   DOMAIN      422    481       29 X 2 AA TANDEM REPEATS OF Q-A.
FT   DOMAIN      674    696       POLY-GLN.
FT   CONFLICT    171    171       N -> T (IN REF. 1).
FT   CONFLICT    302    302       P -> Q (IN REF. 1).
FT   CONFLICT    499    499       N -> T (IN REF. 1).
FT   CONFLICT    751    751       P -> Q (IN REF. 1).
SQ   SEQUENCE   1081 AA;  120308 MW;  275C78721B5415C7 CRC64;

  Query Match             51.2%;  Score 147;  DB 1;  Length 1081;
  Best Local Similarity   51.7%;  Pred. No. 1.8e-06;
  Matches   30;  Conservative   11;  Mismatches   13;  Indels    4;  Gaps    1;

Qy          6 SMATLEKLMKAFESLKSFQ----QQQQQQQQQQQQQQQQQQQQQQQLQPGSTRAAAS 59
              ::|| : : :: : ::  |    |||||||||||||||||||||| : | ||   |:
Db        651 NIATQQNMQQSLQQMQHLQQLKMQQQQQQQQQQQQQQQQQQQQQQQHIYPSSTPGVAN 708
```

```
RESULT 3
TBP_HUMAN
ID   TBP_HUMAN      STANDARD;      PRT;   339 AA.
AC   P20226; Q16845; Q9UC02;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   TATA-box binding protein (TATA-box factor) (TATA binding factor) (TATA
DE   sequence-binding protein) (Transcription initiation factor TFIID TBP
DE   subunit).
GN   TBP OR TFIID OR TF2D.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND DOMAINS.
RX   MEDLINE=90302006; PubMed=2363050;
RA   Peterson M.G., Tanese N., Pugh B.F., Tjian R.;
RT   "Functional domains and upstream activation properties of cloned
RT   human TATA binding protein.";
RL   Science 248:1625-1630(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fibroblast;
RX   MEDLINE=90302010; PubMed=2194289;
RA   Kao C.C., Lieberman P.M., Schmidt M.C., Zhou Q., Pei R., Berk A.J.;
RT   "Cloning of a transcriptionally active human TATA binding factor.";
RL   Science 248:1646-1650(1990).
RN   [3]
RP   SEQUENCE FROM N.A., AND VARIANT 92-GLN--GLN-95 DEL.
RX   MEDLINE=90326195; PubMed=2374612;
RA   Hoffmann A., Sinn E., Yamamoto T., Wang J., Roy A., Horikoshi M.,
RA   Roeder R.G.;
RT   "Highly conserved core domain and unique N terminus with presumptive
RT   regulatory motifs in a human TATA factor (TFIID).";
RL   Nature 346:387-390(1990).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Griffiths C.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   INTERACTION WITH NCOA6.
RX   MEDLINE=20036574; PubMed=10567404;
RA   Lee S.-K., Anzick S.L., Choi J.-E., Bubendorf L., Guan X.-Y.,
RA   Jung Y.-K., Kallioniemi O.P., Kononen J., Trent J.M., Azorsa D.,
RA   Jhun B.-H., Cheong J.H., Lee Y.C., Meltzer P.S., Lee J.W.;
RT   "A nuclear factor ASC-2, as a cancer-amplified transcriptional
RT   coactivator essential for ligand-dependent transactivation by nuclear
RT   receptors in vivo.";
RL   J. Biol. Chem. 274:34283-34293(1999).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 159-337 IN COMPLEX WITH DNA.
RX   MEDLINE=96209823; PubMed=8643494;
RA   Nikolov D.B., Chen H., Halay E.D., Hoffmann A., Roeder R.G.,
RA   Burley S.K.;
```

RT "Crystal structure of a human TATA box-binding protein/TATA element
RT complex.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:4862-4867(1996).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 159-339 IN COMPLEX WITH DNA.
RX MEDLINE=96346176; PubMed=8757291;
RA Juo Z.S., Chiu T.K., Leiberman P.M., Baikalov I., Berk A.J.,
RA Dickerson R.E.;
RT "How proteins recognize the TATA box.";
RL J. Mol. Biol. 261:239-254(1996).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.65 ANGSTROMS) OF 159-337 IN COMPLEX WITH
RP GTF2B AND DNA.
RX MEDLINE=20086817; PubMed=10619841;
RA Tsai F.T.F., Sigler P.B.;
RT "Structural basis of preinitiation complex assembly on human pol II
RT promoters.";
RL EMBO J. 19:25-36(2000).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (2.62 ANGSTROMS) OF 159-339 IN COMPLEX WITH DR1;
RP DRAP1 AND DNA.
RX MEDLINE=21354312; PubMed=11461703;
RA Kamada K., Shu F., Chen H., Malik S., Stelzer G., Roeder R.G.,
RA Meisterernst M., Burley S.K.;
RT "Crystal structure of negative cofactor 2 recognizing the TBP-DNA
RT transcription complex.";
RL Cell 106:71-81(2001).
RN [10]
RP POLYMORPHISM OF POLY-GLN REGION.
RX MEDLINE=99415745; PubMed=10484774;
RA Koide R., Kobayashi S., Shimohata T., Ikeuchi T., Maruyama M.,
RA Saito M., Yamada M., Takahashi H., Tsuji S.;
RT "A neurological disease caused by an expanded CAG trinucleotide repeat
RT in the TATA-binding protein gene: a new polyglutamine disease?";
RL Hum. Mol. Genet. 8:2047-2053(1999).
RN [11]
RP POLYMORPHISM OF POLY-GLN REGION.
RX MEDLINE=21214723; PubMed=11313753;
RA Zuhlke C., Hellenbroich Y., Dalski A., Kononowa N., Hagenah J.,
RA Vieregge P., Riess O., Klein C., Schwinger E.;
RT "Different types of repeat expansion in the TATA-binding protein gene
RT are associated with a new form of inherited ataxia.";
RL Eur. J. Hum. Genet. 9:160-164(2001).
RN [12]
RP POLYMORPHISM OF POLY-GLN REGION.
RX MEDLINE=21341926; PubMed=11448935;
RA Nakamura K., Jeong S.-Y., Uchihara T., Anno M., Nagashima K.,
RA Nagashima T., Ikeda S.-I., Tsuji S., Kanazawa I.;
RT "SCA17, a novel autosomal dominant cerebellar ataxia caused by an
RT expanded polyglutamine in TATA-binding protein.";
RL Hum. Mol. Genet. 10:1441-1448(2001).
RN [13]
RP POLYMORPHISM OF POLY-GLN REGION.
RX MEDLINE=21937712; PubMed=11939898;
RA Silveira I., Miranda C., Guimaraes L., Moreira M.-C., Alonso I.,
RA Mendonca P., Ferro A., Pinto-Basto J., Coelho J., Ferreirinha F.,
RA Poirier J., Parreira E., Vale J., Januario C., Barbot C., Tuna A.,

RA Barros J., Koide R., Tsuji S., Holmes S.E., Margolis R.L., Jardim L.,
RA Pandolfo M., Coutinho P., Sequeiros J.;
RT "Trinucleotide repeats in 202 families with ataxia: a small expanded
RT (CAG)n allele at the SCA17 locus.";
RL Arch. Neurol. 59:623-629(2002).
CC -!- FUNCTION: General transcription factor that functions at the
CC core of the DNA-binding multiprotein factor TFIID. Binding of
CC TFIID to the TATA box is the initial transcriptional step of the
CC pre-initiation complex (PIC), playing a role in the activation of
CC eukaryotic genes transcribed by RNA polymerase II.
CC -!- SUBUNIT: Belongs to the TFIID complex together with the TBP-
CC associated factors (TAFs). Binds DNA as monomer. Interacts with
CC TAFs, TFIIA, TFIIB, NCOA6, DRAP1 and DR1.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- POLYMORPHISM: The poly-Gln region of TBP is highly polymorphic (25
CC to 42 repeats) in normal individuals and is expanded to about 47-
CC 63 repeats in SCA17 patients. Longer expansions may result in
CC earlier onset and more severe clinical manifestations of the
CC disease.
CC -!- DISEASE: Defects in TBP are the cause of spinocerebellar ataxia
CC type 17 (SCA17) [MIM:607136]. SCA17 is a rare autosomal dominant
CC neurodegenerative disease, characterized by gait ataxia and
CC dementia, progressing over several decades to include
CC bradykinesia, dysmetria, dysdiadochokinesis, hyperreflexia and
CC paucity of movement.
CC -!- SIMILARITY: Belongs to the TBP family.
CC --

CC --
DR EMBL; M55654; AAA36731.1; -.
DR EMBL; M34960; AAC03409.1; -.
DR EMBL; X54993; CAA38736.1; -.
DR EMBL; AL031259; CAA20286.1; -.
DR PIR; A34830; TWHU2D.
DR PDB; 1CDW; 23-DEC-96.
DR PDB; 1C9B; 10-JAN-00.
DR PDB; 1JFI; 11-JUL-01.
DR PDB; 1TGH; 01-AUG-96.
DR TRANSFAC; T00794; -.
DR Genew; HGNC:11588; TBP.
DR MIM; 600075; -.
DR MIM; 607136; -.
DR GO; GO:0005669; C:transcription factor TFIID complex; TAS.
DR GO; GO:0016251; F:general RNA polymerase II transcription fac. . .; TAS.
DR GO; GO:0006367; P:transcription initiation from Pol II promoter; TAS.
DR InterPro; IPR000814; TFIID.
DR Pfam; PF00352; TBP; 2.
DR PRINTS; PR00686; TIFACTORIID.
DR PROSITE; PS00351; TFIID; 2.
KW Transcription; Nuclear protein; DNA-binding; Repeat; Polymorphism;
KW Triplet repeat expansion; Disease mutation; 3D-structure.

```
FT   REPEAT       165    241       1.
FT   REPEAT       255    332       2.
FT   DOMAIN        55     95       POLY-GLN.
FT   VARIANT       92     95       Missing.
FT                                  /FTId=VAR_016987.
FT   CONFLICT     187    187       A -> R (IN REF. 2).
FT   STRAND       164    173
FT   HELIX        180    186
FT   TURN         188    189
FT   STRAND       190    192
FT   TURN         194    196
FT   STRAND       199    205
FT   TURN         206    207
FT   STRAND       208    214
FT   TURN         215    216
FT   STRAND       218    222
FT   HELIX        227    244
FT   TURN         245    245
FT   STRAND       251    263
FT   STRAND       268    268
FT   HELIX        270    276
FT   TURN         277    280
FT   STRAND       281    282
FT   TURN         285    287
FT   STRAND       291    295
FT   TURN         296    299
FT   STRAND       300    304
FT   TURN         306    307
FT   STRAND       309    313
FT   HELIX        318    333
FT   TURN         334    335
FT   STRAND       336    336

  Query Match             50.5%;  Score 145;  DB 1;  Length 339;
  Best Local Similarity   59.3%;  Pred. No. 9.7e-07;
  Matches   32;  Conservative    7;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          6 SMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQPGSTRAAAS 59
              |:: ||:  :  :  :  |||||||||||||||||||||||| |     |||:
Db         48 SLSILEEQQRQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQAVAAAA 101


RESULT 4
RUN2_MOUSE
ID   RUN2_MOUSE     STANDARD;      PRT;   607 AA.
AC   Q08775; O35183; Q08776; Q9JLN0; Q9QUQ6; Q9QY29; Q9R0U4; Q9Z2J7;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Runt-related transcription factor 2 (Core-binding factor, alpha 1
DE   subunit) (CBF-alpha 1) (Acute myeloid leukemia 3 protein) (Oncogene
DE   AML-3) (Polyomavirus enhancer binding protein 2 alpha A subunit)
DE   (PEBP2-alpha A) (PEA2-alpha A) (SL3-3 enhancer factor 1 alpha A
DE   subunit) (SL3/AKV core-binding factor alpha A subunit) (Osteoblast-
DE   specific transcription factor 2) (OSF-2).
GN   RUNX2 OR CBFA1 OR AML3 OR PEBP2A OR OSF2.
OS   Mus musculus (Mouse).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
RX   MEDLINE=93342088; PubMed=8341710;
RA   Ogawa E., Maruyama M., Kagoshima H., Inuzuka M., Lu J., Satake M.,
RA   Shigesada K., Ito Y.;
RT   "PEBP2/PEA2 represents a family of transcription factors homologous to
RT   the products of the Drosophila runt gene and the human AML1 gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:6859-6863(1993).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   STRAIN=C57BL/6;
RC   TISSUE=Osteoblast;
RX   MEDLINE=97325750; PubMed=9182762;
RA   Ducy P., Zhang R., Geoffroy V., Ridall A.L., Karsenty G.;
RT   "Osf2/Cbfa1: a transcriptional activator of osteoblast
RT   differentiation.";
RL   Cell 89:747-754(1997).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4; 5; 6; 7; 8 AND 9).
RC   STRAIN=CD2-MYC;
RX   MEDLINE=97385157; PubMed=9238031;
RA   Stewart M., Terry A., Hu M., O'Hara M., Blyth K., Baxter E.,
RA   Cameron E., Onions D.E., Neil J.C.;
RT   "Proviral insertions induce the expression of bone-specific isoforms
RT   of PEBP2alphaA (CBFA1): evidence for a new myc collaborating
RT   oncogene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:8646-8651(1997).
RN   [4]
RP   PARTIAL SEQUENCE FROM N.A. (ISOFORMS 2 AND 6), AND ALTERNATIVE
RP   SPLICING.
RX   MEDLINE=98322266; PubMed=9651525;
RA   Xiao Z.S., Thomas R., Hinson T.K., Quarles L.D.;
RT   "Genomic structure and isoform expression of the mouse, rat and human
RT   Cbfa1/Osf2 transcription factor.";
RL   Gene 214:187-197(1998).
RN   [5]
RP   SEQUENCE OF 1-98 FROM N.A. (ISOFORMS 1 AND 2).
RX   MEDLINE=99453726; PubMed=10524201;
RA   Fujiwara M., Tagashira S., Harada H., Ogawa S., Katsumata T.,
RA   Nakatsuka M., Komori T., Takada H.;
RT   "Isolation and characterization of the distal promoter region of mouse
RT   Cbfa1.";
RL   Biochim. Biophys. Acta 1446:265-272(1999).
RN   [6]
RP   SEQUENCE OF 263-277 AND 305-319.
RX   MEDLINE=93242761; PubMed=8386878;
RA   Ogawa E., Inuzuka M., Maruyama M., Satake M., Naito-Fujimoto M.,
RA   Ito Y., Shigesada K.;
RT   "Molecular cloning and characterization of PEBP2 beta, the
RT   heterodimeric partner of a novel Drosophila runt-related DNA binding
RT   protein PEBP2 alpha.";
RL   Virology 194:314-331(1993).
RN   [7]
RP   SEQUENCE OF 1-35 FROM N.A.
```

```
RC   STRAIN=129;
RA   Chi X.-Z., Bae S.-C.;
RT   "Analysis of the two PEBP2aA/cbfa1 promoter regions.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   FUNCTION.
RX   MEDLINE=97325751; PubMed=9182763;
RA   Komori T., Yagi H., Nomura S., Yamaguchi A., Sasaki K., Deguchi K.,
RA   Shimizu Y., Bronson R.T., Gao Y.-H., Inada M., Sato M., Okamoto R.,
RA   Kitamura Y., Yoshiki S., Kishimoto T.;
RT   "Targeted disruption of Cbfa1 results in a complete lack of bone
RT   formation owing to maturational arrest of osteoblasts.";
RL   Cell 89:755-764(1997).
RN   [9]
RP   PHOSPHORYLATION.
RX   MEDLINE=20127938; PubMed=10660618;
RA   Xiao G., Jiang D., Thomas P., Benson M.D., Guan K., Karsenty G.,
RA   Franceschi R.T.;
RT   "MAPK pathways activate and phosphorylate the osteoblast-specific
RT   transcription factor, Cbfa1.";
RL   J. Biol. Chem. 275:4453-4459(2000).
CC   -!- FUNCTION: Transcription factor involved in osteoblastic
CC       differentiation and skeletal morphogenesis. Essential for the
CC       maturation of osteoblasts and both intramembranous and
CC       endochondral ossification. Cbf binds to the core site, 5'-
CC       PYGPYGGT-3', of a number of enhancers and promoters, including
CC       murine leukemia virus, polyomavirus enhancer, T-cell receptor
CC       enhancers, osteocalcin, osteopontin, bone sialoprotein, alpha 1(I)
CC       collagen, LCK, IL-3 and GM-CSF promoters.
CC   -!- SUBUNIT: Heterodimer of an alpha and a beta subunit. The alpha
CC       subunit binds DNA as a monomer and through the Runt domain. DNA-
CC       binding is increased by heterodimerization.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=9;
CC         Comment=Additional isoforms seem to exist;
CC       Name=1;
CC         IsoId=Q08775-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q08775-2; Sequence=VSP_005941;
CC       Name=3; Synonyms=PEBP2-alpha A1;
CC         IsoId=Q08775-3; Sequence=VSP_005940, VSP_005942;
CC       Name=4; Synonyms=PEBP2-alpha A2;
CC         IsoId=Q08775-4; Sequence=VSP_005940, VSP_005942, VSP_005944,
CC                                  VSP_005945;
CC       Name=5; Synonyms=G1;
CC         IsoId=Q08775-5; Sequence=VSP_005939;
CC       Name=6; Synonyms=G2;
CC         IsoId=Q08775-6; Sequence=VSP_005939, VSP_005943;
CC       Name=7; Synonyms=U1;
CC         IsoId=Q08775-7; Sequence=VSP_005939, VSP_005946, VSP_005948;
CC       Name=8; Synonyms=Y1;
CC         IsoId=Q08775-8; Sequence=VSP_005939, VSP_005947;
CC       Name=9; Synonyms=Y2;
CC         IsoId=Q08775-9; Sequence=VSP_005939, VSP_005943, VSP_005947;
CC   -!- TISSUE SPECIFICITY: Found in thymus and testis, T cell lines but
CC       not in B-cell lines. Isoform 2 is exclusively found in bone,
```

```
CC       particularly in osteoblasts; isoforms 3 and 4 are expressed in T-
CC       cell lines; isoforms 5, 6, 7, 8 and 9 can be found in osteoblasts
CC       and osteosarcoma cell lines.
CC   -!- DEVELOPMENTAL STAGE: Expression occurs early during skeletal
CC       development and is restricted to cells of the mesenchymal
CC       condensations and of the osteoblast lineage. Expression of isoform
CC       2 in the embryo reaches a peak at 12.5 dpc.
CC   -!- DOMAIN: A proline/serine/threonine rich region at the C-terminus
CC       is necessary for transcriptional activation of target genes and
CC       contains the phosphorylation sites.
CC   -!- PTM: Phosphorylated; probably by MAP kinases (MAPK).
CC   -!- SIMILARITY: Contains 1 Runt domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D14636; BAA03485.1; -.
DR   EMBL; D14637; BAA03486.1; -.
DR   EMBL; AF010284; AAB65409.1; -.
DR   EMBL; AF005936; AAB82419.1; -.
DR   EMBL; AF053948; AAC77440.1; -.
DR   EMBL; AF053951; AAC78623.1; -.
DR   EMBL; AF053956; AAC78626.1; -.
DR   EMBL; AF134836; AAF22568.1; -.
DR   EMBL; AF134836; AAF22569.1; -.
DR   EMBL; AB013129; BAA85345.1; -.
DR   EMBL; AB013129; BAA85346.1; -.
DR   EMBL; AF155360; AAF73290.1; -.
DR   PIR; A48233; A48233.
DR   HSSP; O60472; 1CMO.
DR   TRANSFAC; T01062; -.
DR   TRANSFAC; T01063; -.
DR   MGD; MGI:99829; Runx2.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   GO; GO:0045944; P:positive regulation of transcription from P. . .; IDA.
DR   InterPro; IPR000040; AML1_Runt.
DR   InterPro; IPR008967; P53-like.
DR   Pfam; PF00853; Runt; 1.
DR   PRINTS; PR00967; ONCOGENEAML1.
KW   Transcription regulation; DNA-binding; Nuclear protein; ATP-binding;
KW   Alternative splicing; Phosphorylation.
FT   DOMAIN      187    314       RUNT.
FT   DOMAIN      323    607       PRO/SER/THR-RICH.
FT   DOMAIN      128    156       POLY-GLN.
FT   DOMAIN      158    175       POLY-ALA.
FT   NP_BIND     275    282       ATP (POTENTIAL).
FT   VARSPLIC      1     79       Missing (in isoform 5, isoform 6, isoform
FT                                7, isoform 8 and isoform 9).
FT                                /FTId=VSP_005939.
FT   VARSPLIC      1     98       MLHSPHKQPQNHKCGANFLQEDCKKALAFKWLISAGHYQPP
FT                                RPTESVSALTTVHAGIFKAASSIYNRGHKFYLEKKGGTMAS
```

```
FT                                NSLFSAVTPCQQSFFW -> MRIPV (in isoform 3
FT                                and isoform 4).
FT                                /FTId=VSP_005940.
FT   VARSPLIC     47     57       Missing (in isoform 2).
FT                                /FTId=VSP_005941.
FT   VARSPLIC    156    156       Missing (in isoform 3 and isoform 4).
FT                                /FTId=VSP_005942.
FT   VARSPLIC    316    373       Missing (in isoform 6 and isoform 9).
FT                                /FTId=VSP_005943.
FT   VARSPLIC    399    400       PS -> LS (in isoform 4).
FT                                /FTId=VSP_005944.
FT   VARSPLIC    401    607       Missing (in isoform 4).
FT                                /FTId=VSP_005945.
FT   VARSPLIC    427    439       DDDTATSDFCLWP -> GFCGTTTTTTTKL (in
FT                                isoform 7).
FT                                /FTId=VSP_005946.

  Query Match             49.7%;  Score 142.5;  DB 1;  Length 607;
  Best Local Similarity   58.2%;  Pred. No. 2.8e-06;
  Matches   32;  Conservative    7;  Mismatches   11;  Indels    5;  Gaps    1;

Qy          5 GSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQPGSTRAAAS 59
              | |: :  :: | :     ||||||||||||||||||||||||| |  :  |||:
Db        116 GKMSDVSPVVAAQQ-----QQQQQQQQQQQQQQQQQQQQQQQQQQEAAAAAAAA 165


RESULT 5
SSN6_YEAST
ID   SSN6_YEAST     STANDARD;      PRT;   966 AA.
AC   P14922;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   Glucose repression mediator protein.
GN   SSN6 OR CYC8 OR YBR112C OR YBR0908.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89211964; PubMed=2854095;
RA   Trumbly R.J.;
RT   "Cloning and characterization of the CYC8 gene mediating glucose
RT   repression in yeast.";
RL   Gene 73:97-111(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88065502; PubMed=3316983;
RA   Schultz J., Carlson M.;
RT   "Molecular analysis of SSN6, a gene functionally related to the SNF1
RT   protein kinase of Saccharomyces cerevisiae.";
RL   Mol. Cell. Biol. 7:3637-3645(1987).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=92327848; PubMed=1626431;
```

```
RA   Mannhaupt G., Stucka R., Ehnle S., Vetter I., Feldmann H.;
RT   "Molecular analysis of yeast chromosome II between CMD1 and LYS2: the
RT   excision repair gene RAD16 located in this region belongs to a novel
RT   group of double-finger proteins.";
RL   Yeast 8:397-408(1992).
RN   [4]
RP   TPR REPEATS.
RX   MEDLINE=90124639; PubMed=2404612;
RA   Sikorski R.S., Boguski M.S., Goebl M., Hieter P.A.;
RT   "A repeating amino acid motif in CDC23 defines a family of proteins
RT   and a new relationship among genes required for mitosis and RNA
RT   synthesis.";
RL   Cell 60:307-317(1990).
CC   -!- FUNCTION: IT IS INVOLVED IN REPRESSION BY A1-ALPHA2 AND ALPHA2 AND
CC       IN OTHER SYSTEMS AS A GENERAL REPRESSOR OF TRANSCRIPTION. THIS
CC       PROTEIN HAS NO OBVIOUS DNA-BINDING DOMAINS. IT MIGHT NOT INTERACT
CC       DIRECTLY WITH DNA BUT WITH DNA-BOUND PROTEINS.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: Contains 10 TPR repeats.
CC   -!- SIMILARITY: TO YEAST GAL1 AND CCR4.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M23440; AAA34545.1; -.
DR   EMBL; M17826; AAA35103.1; -.
DR   EMBL; X66247; CAA46973.1; -.
DR   EMBL; X78993; CAA55615.1; -.
DR   EMBL; Z35981; CAA85069.1; -.
DR   PIR; S25365; S25365.
DR   GermOnline; 138655; -.
DR   TRANSFAC; T03687; -.
DR   SGD; S0000316; CYC8.
DR   GO; GO:0005634; C:nucleus; IPI.
DR   GO; GO:0016565; F:general transcriptional repressor activity; IDA.
DR   GO; GO:0003713; F:transcription co-activator activity; IDA.
DR   GO; GO:0016481; P:negative regulation of transcription; IDA.
DR   InterPro; IPR008941; TPR-like.
DR   InterPro; IPR001440; TPR.
DR   Pfam; PF00515; TPR; 10.
DR   SMART; SM00028; TPR; 9.
KW   Transcription regulation; Repressor; Repeat; TPR repeat;
KW   Nuclear protein.
FT   DOMAIN       15     30       POLY-GLN.
FT   REPEAT       46     79       TPR 1.
FT   REPEAT       80    113       TPR 2.
FT   REPEAT      114    147       TPR 3.
FT   REPEAT      150    183       TPR 4.
FT   REPEAT      187    220       TPR 5.
FT   REPEAT      224    257       TPR 6.
FT   REPEAT      258    291       TPR 7.
FT   REPEAT      296    329       TPR 8.
```

```
FT   REPEAT       330     363       TPR 9.
FT   REPEAT       364     398       TPR 10.
FT   DOMAIN       493     556       30 X 2 AA TANDEM REPEATS OF Q-A.
FT   DOMAIN       557     587       POLY-GLN.
FT   CONFLICT     547     547       K -> Q (IN REF. 3).
SQ   SEQUENCE   966 AA;  107202 MW;  84B509CF3208C5C0 CRC64;

  Query Match             49.1%;  Score 141;  DB 1;  Length 966;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-06;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         24 QQQQQQQQQQQQQQQQQQQQQQQQQLQP 51
              ||||||||||||||||||||||||||||
Db        563 QQQQQQQQQQQQQQQQQQQQQQQQQLQP 590


RESULT 6
HMAA_DROME
ID   HMAA_DROME     STANDARD;      PRT;   590 AA.
AC   P29555; Q9VER1;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Homeobox protein abdominal-A.
GN   ABD-A OR CG10325.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Canton-S;
RX   MEDLINE=95396803; PubMed=7667301;
RA   Martin C.H., Mayeda C.A., Davis C.A., Ericsson C.L., Knafels J.D.,
RA   Mathog D.R., Celniker S.E., Lewis E.B., Palazzolo M.J.;
RT   "Complete sequence of the bithorax complex of Drosophila.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:8398-8402(1995).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM ABD-A1).
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
```

```
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [3]
RP   SEQUENCE OF 261-590 FROM N.A.
RX   MEDLINE=91071585; PubMed=1979297;
RA   Karch F., Bender W., Weiffenbach B.;
RT   "abdA expression in Drosophila embryos.";
RL   Genes Dev. 4:1573-1587(1990).
CC   -!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
CC       A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
CC       SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
CC       REQUIRED FOR SEGMENTAL IDENTITY OF THE SECOND THROUGH EIGHTH
CC       ABDOMINAL SEGMENTS. ONCE A PATTERN OF ABD-A EXPRESSION IS TURNED
CC       ON IN A GIVEN PARASEGMENT, IT REMAINS ON THE MORE POSTERIOR
CC       PARASEGMENT, SO THAT THE COMPLEX PATTERN OF EXPRESSION IS BUILT UP
CC       IN THE SUCCESSIVE PARASEGMENTS. APPEARS TO REPRESS EXPRESSION OF
CC       UBX WHENEVER THEY APPEAR IN THE SAME CELL, BUT ABD-A IS REPRESSED
CC       BY ABDB ONLY IN THE EIGHT AND NINTH ABDOMINAL SEGMENTS.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Abd-A2;
CC         IsoId=P29555-1; Sequence=Displayed;
CC       Name=Abd-A1;
CC         IsoId=P29555-2; Sequence=VSP_002394;
CC   -!- SIMILARITY: Belongs to the Antp homeobox family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U31961; AAA84405.1; -.
```

```
DR   EMBL; U31961; AAA84406.1; -.
DR   EMBL; X54453; CAA38321.1; -.
DR   EMBL; AE003715; AAF55359.1; -.
DR   PIR; A35915; A35915.
DR   HSSP; P02833; 9ANT.
DR   TRANSFAC; T01992; -.
DR   FlyBase; FBgn0000014; abd-A.
DR   GO; GO:0007438; P:oenocyte development; IMP.
DR   InterPro; IPR001827; Antennapedia.
DR   InterPro; IPR001356; Homeobox.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00025; ANTENNAPEDIA.
DR   PRINTS; PR00024; HOMEOBOX.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS00032; ANTENNAPEDIA; FALSE_NEG.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW   Alternative splicing.
FT   DOMAIN       35     50       POLY-ALA.
FT   DOMAIN       51    119       SER-RICH.
FT   DOMAIN      136    139       POLY-GLN (OPA-REPEAT).
FT   DOMAIN      144    147       POLY-GLN (OPA-REPEAT).
FT   DOMAIN      160    165       POLY-GLN (OPA-REPEAT).
FT   DOMAIN      172    177       POLY-ALA.
FT   DOMAIN      240    250       POLY-ALA.
FT   SITE        368    373       ANTP-TYPE HEXAPEPTIDE.
FT   DNA_BIND    398    457       HOMEOBOX.
FT   DOMAIN      425    428       POLY-ARG.
FT   DOMAIN      491    518       POLY-GLN (OPA-REPEAT).
FT   VARSPLIC      1    260       Missing (in isoform Abd-A1).
FT                                /FTId=VSP_002394.
SQ   SEQUENCE   590 AA;  62409 MW;  FF080CC2D71ECA82 CRC64;

  Query Match             49.0%;  Score 140.5;  DB 1;  Length 590;
  Best Local Similarity   70.5%;  Pred. No. 4.1e-06;
  Matches   31;  Conservative    4;  Mismatches    6;  Indels    3;  Gaps    1;

Qy         11 EKLMKAFESLKSFQQQ---QQQQQQQQQQQQQQQQQQQQQLQP 51
              :: ||| |::|| ||    |||||||||||||||| |||| ||
Db        474 QEKMKAQETMKSAQQNKQVQQQQQQQQQQQQQQQQQHQQQQQQP 517


RESULT 7
NCO6_HUMAN
ID   NCO6_HUMAN     STANDARD;      PRT;  2063 AA.
AC   Q14686; Q9NTZ9; Q9UH74; Q9UK86;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Nuclear receptor coactivator 6 (Amplified in breast cancer-3 protein)
DE   (Cancer-amplified transcriptional coactivator ASC-2) (Activating
DE   signal cointegrator-2) (ASC-2) (Peroxisome proliferator-activated
DE   receptor-interacting protein) (PPAR-interacting protein) (PRIP)
DE   (Nuclear receptor-activating protein, 250 kDa) (Nuclear receptor
DE   coactivator RAP250) (NRC RAP250) (Thyroid hormone receptor-binding
```

```
DE   protein).
GN   NCOA6 OR AIB3 OR RAP250 OR TRBP OR KIAA0181.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND INTERACTION WITH CREBBP; NCOA1; GTF2A; TBP;
RP   RXRA; ESR1; RARA AND THRA.
RX   MEDLINE=20036574; PubMed=10567404;
RA   Lee S.-K., Anzick S.L., Choi J.-E., Bubendorf L., Guan X.-Y.,
RA   Jung Y.-K., Kallioniemi O.P., Kononen J., Trent J.M., Azorsa D.,
RA   Jhun B.-H., Cheong J.H., Lee Y.C., Meltzer P.S., Lee J.W.;
RT   "A nuclear factor ASC-2, as a cancer-amplified transcriptional
RT   coactivator essential for ligand-dependent transactivation by nuclear
RT   receptors in vivo.";
RL   J. Biol. Chem. 274:34283-34293(1999).
RN   [2]
RP   SEQUENCE FROM N.A., HOMODIMERIZATION, AND INTERACTION WITH CREBBP;
RP   RXRA; ESR1; NR3C1; RARA; VDR AND THRA.
RX   MEDLINE=20325329; PubMed=10866662;
RA   Mahajan M.A., Samuels H.H.;
RT   "A new family of nuclear receptor coregulators that integrates nuclear
RT   receptor signaling through CBP.";
RL   Mol. Cell. Biol. 20:5048-5063(2000).
RN   [3]
RP   SEQUENCE FROM N.A., AND INTERACTION WITH PPARA; PPARG; ESR1; ESR2 AND
RP   THR.
RC   TISSUE=Testis;
RX   MEDLINE=20148724; PubMed=10681503;
RA   Caira F., Antonson P., Pelto-Huikko M., Treuter E., Gustafsson J.-A.;
RT   "Cloning and characterization of RAP250, a nuclear receptor
RT   coactivator.";
RL   J. Biol. Chem. 275:5308-5317(2000).
RN   [4]
RP   SEQUENCE FROM N.A., PHOSPHORYLATION BY PRKDC, AND INTERACTION WITH
RP   THR; RAR; EP300 AND CRSP3.
RC   TISSUE=Lymphocytes;
RX   MEDLINE=20283976; PubMed=10823961;
RA   Ko L., Cardona G.R., Chin W.W.;
RT   "Thyroid hormone receptor-binding protein, an LXXLL motif-containing
RT   protein, functions as a general coactivator.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:6212-6217(2000).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=96281124; PubMed=8724849;
RA   Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
RT   "Prediction of the coding sequences of unidentified human genes. V.
RT   The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
RT   analysis of cDNA clones from human cell line KG-1.";
RL   DNA Res. 3:17-24(1996).
RN   [6]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21638749; PubMed=11780052;
RA   Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA   Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
```

RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Lehvaslaiho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [7]
RP INTERACTION WITH NCOA6IP.
RX MEDLINE=21417756; PubMed=11517327;
RA Zhu Y.-J., Qi C., Cao W.-Q., Yeldandi A.V., Rao M.S., Reddy J.K.;
RT "Cloning and characterization of PIMT, a protein with a
RT methyltransferase domain, which interacts with and enhances nuclear
RT receptor coactivator PRIP function.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:10380-10385(2001).
RN [8]
RP INTERACTION WITH RBM14.
RX MEDLINE=21423995; PubMed=11443112;
RA Iwasaki T., Chin W.W., Ko L.;
RT "Identification and characterization of RRM-containing coactivator
RT activator (CoAA) as TRBP-interacting protein, and its splice variant
RT as a coactivator modulator (CoAM).";
RL J. Biol. Chem. 276:33375-33383(2001).
RN [9]
RP INTERACTION WITH HRMT1L1.
RX MEDLINE=22151129; PubMed=12039952;
RA Qi C., Chang J., Zhu Y., Yeldandi A.V., Rao S.M., Zhu Y.-J.;
RT "Identification of protein arginine methyltransferase 2 as a
RT coactivator for estrogen receptor alpha.";
RL J. Biol. Chem. 277:28624-28630(2002).
RN [10]
RP INTERACTION WITH MLL3 AND THE ASCOM COMPLEX.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=22371496; PubMed=12482968;
RA Goo Y.-H., Sohn Y.C., Kim D.-H., Kim S.-W., Kang M.-J., Jung D.-J.,
RA Kwak E., Barlev N.A., Berger S.L., Chow V.T., Roeder R.G.,
RA Azorsa D.O., Meltzer P.S., Suh P.-G., Song E.J., Lee K.-J., Lee Y.C.,
RA Lee J.W.;
RT "Activating signal cointegrator 2 belongs to a novel steady-state

```
RT   complex that contains a subset of trithorax group proteins.";
RL   Mol. Cell. Biol. 23:140-149(2003).
RN   [11]
RP   MUTAGENESIS OF 883-THR--GLU-894, AND PHOSPHORYLATION.
RX   MEDLINE=21635582; PubMed=11773444;
RA   Ko L., Cardona G.R., Iwasaki T., Bramlett K.S., Burris T.P.,
RA   Chin W.W.;
RT   "Ser-884 adjacent to the LXXLL motif of coactivator TRBP defines
RT   selectivity for ERs and TRs.";
RL   Mol. Endocrinol. 16:128-140(2002).
CC   -!- FUNCTION: Nuclear receptor coactivator that directly binds nuclear
CC       receptors and stimulates the transcriptional activities in a
CC       hormone-dependent fashion. Coactivates expression in an agonist-
CC       and AF2-dependent manner. Involved in the coactivation of
CC       different nuclear receptors, such as for steroids (GR and ERs),
CC       retinoids (RARs and RXRs), thyroid hormone (TRs), vitamin D3 (VDR)
CC       and prostanoids (PPARs). Probably functions as a general
CC       coactivator, rather than just a nuclear receptor coactivator. May
CC       also be involved in the coactivation of the NF-kappa-B pathway.
CC       May coactivate expression via a remodeling of chromatin and its
CC       interaction with histone acetyltransferase proteins.
CC   -!- SUBUNIT: Monomer and homodimer. Interacts with RNPC2 (By
CC       similarity). Interacts in vitro with the basal transcription
CC       factors GTF2A and TBP, suggesting an autonomous transactivation
CC       function. Interacts with NCOA1, CRSP3, RBM14, the histone
CC       acetyltransferases EP300 and CREBBP, and with the
CC       methyltransferases NCOA6IP and HRMT1L1/PRMT2. Belongs to the
CC       ASC-2/NCOA6 complex (ASCOM), which contains ASC-2/NCOA6, the
CC       retinoblastoma-binding protein RBQ-3/ RBBP5, alpha- and beta-
CC       tubulins, the trithorax group proteins MLL2 and MLL3, and
CC       ASH2/ASCL2.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: Ubiquitous. Highly expressed in brain,
CC       prostate, testis and ovary; weakly expressed in lung, thymus and
CC       small intestine.
CC   -!- DOMAIN: Contains two Leu-Xaa-Xaa-Leu-Leu (LXXLL) motifs. Only
CC       motif 1 is essential for the association with nuclear receptors,
CC       while adjacent Ser-884 displays selectivity for nuclear receptors.
CC   -!- PTM: Phosphorylated by PRKDC.
CC   -!- PTM: Phosphorylation on Ser-884 leads to a strong decrease in
CC       binding to ESR1 and ESR2.
CC   -!- MISCELLANEOUS: Frequently amplified or ovexpressed in colon,
CC       breast and lung cancers.
CC   -!- CAUTION: Ref.1 (AAF16403) sequence differs from that shown due to
CC       a frameshift in position 88.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF177388; AAF13595.1; -.
DR   EMBL; AF208227; AAF16403.1; ALT_FRAME.
DR   EMBL; AF245115; AAF78480.1; -.
```

```
DR   EMBL; AF128458; AAF37003.1; -.
DR   EMBL; AF171667; AAF71829.1; -.
DR   EMBL; D80003; BAA11498.2; ALT_INIT.
DR   EMBL; AL109824; CAB92721.1; -.
DR   Genew; HGNC:15936; NCOA6.
DR   MIM; 605299; -.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0005667; C:transcription factor complex; TAS.
DR   GO; GO:0003682; F:chromatin binding; ISS.
DR   GO; GO:0030331; F:estrogen receptor binding; TAS.
DR   GO; GO:0046965; F:retinoid X receptor binding; TAS.
DR   GO; GO:0046966; F:thyroid hormone receptor binding; IDA.
DR   GO; GO:0003713; F:transcription co-activator activity; IDA.
DR   GO; GO:0016563; F:transcriptional activator activity; TAS.
DR   GO; GO:0007420; P:brain development; ISS.
DR   GO; GO:0001701; P:embryonic development (sensu Mammalia); ISS.
DR   GO; GO:0007507; P:heart development; ISS.
DR   GO; GO:0030099; P:myeloid blood cell differentiation; IDA.

  Query Match             48.1%;  Score 138;  DB 1;  Length 2063;
  Best Local Similarity   57.1%;  Pred. No. 2.2e-05;
  Matches   32;  Conservative    3;  Mismatches   13;  Indels    8;  Gaps    1;

Qy          3 PRGSMATLEKLMKAFE---------SLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQ 50
              | ||:|     |:           :    |||||||||||||||||||||||||||
Db        232 PSGSLAPPHHPMQPVSVNRQMNPANFPQLQQQQQQQQQQQQQQQQQQQQQQQQQLQ 287


RESULT 8
MJD1_HUMAN
ID   MJD1_HUMAN     STANDARD;      PRT;   376 AA.
AC   P54252; O15284; O15285; O15286; Q8N189; Q96TC3; Q96TC4; Q9H3N0;
DT   01-OCT-1996 (Rel. 34, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Machado-Joseph disease protein 1 (Ataxin-3) (Spinocerebellar ataxia
DE   type 3 protein).
GN   MJD OR MJD1 OR SCA3 OR ATX3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANT MJD1A.
RC   TISSUE=Brain;
RX   MEDLINE=95179166; PubMed=7874163;
RA   Kawaguchi Y., Okamoto T., Taniwaki M., Aizawa M., Inoue M.,
RA   Katayama S., Kawakami H., Nakamura S., Nishimura M., Akiguchi I.,
RA   Kimura J., Narumiya S., Kakizuka A.;
RT   "CAG expansions in a novel gene for Machado-Joseph disease at
RT   chromosome 14q32.1.";
RL   Nat. Genet. 8:221-228(1994).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANTS VAL-212;
RP   AND MJD1A.
RX   MEDLINE=97418757; PubMed=9274833;
RA   Goto J., Watanabe M., Ichikawa Y., Yee S.-B., Ihara N., Endo K.,
```

RA Igarashi S., Takiyama Y., Gaspar C., Maciel P., Tsuji S.,
RA Rouleau G.A., Kanazawa I.;
RT "Machado-Joseph disease gene products carrying different carboxyl
RT termini.";
RL Neurosci. Res. 28:373-377(1997).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3), AND VARIANT MJD1A.
RX MEDLINE=21342815; PubMed=11450850;
RA Ichikawa Y., Goto J., Hattori M., Toyoda A., Ishii K., Jeong S.-Y.,
RA Hashida H., Masuda N., Ogata K., Kasai F., Hirai M., Maciel P.,
RA Rouleau G.A., Sakaki Y., Kanazawa I.;
RT "The genomic structure and expression of MJD, the Machado-Joseph
RT disease gene.";
RL J. Hum. Genet. 46:413-422(2001).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANT VAL-212.
RC TISSUE=Breast;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [5]
RP SUBCELLULAR LOCATION.
RX MEDLINE=98248424; PubMed=9580663;
RA Tait D., Riccio M., Sittler A., Scherzinger E., Santi S., Ognibene A.,
RA Maraldi N.M., Lehrach H., Wanker E.E.;
RT "Ataxin-3 is transported into the nucleus and associates with the
RT nuclear matrix.";
RL Hum. Mol. Genet. 7:991-997(1998).
RN [6]
RP FUNCTION.
RX MEDLINE=22323318; PubMed=12297501;
RA Li F., Macfarlan T., Pittman R.N., Chakravarti D.;
RT "Ataxin-3 is a histone-binding protein with two independent
RT transcriptional corepressor activities.";
RL J. Biol. Chem. 277:45004-45012(2002).
RN [7]
RP 3D-STRUCTURE MODELING.
RX MEDLINE=22374627; PubMed=12486728;
RA Albrecht M., Hoffmann D., Evert B.O., Schmitt I., Wuellner U.,

```
RA   Lengauer T.;
RT   "Structural modeling of ataxin-3 reveals distant homology to
RT   adaptins.";
RL   Proteins 50:355-370(2003).
CC   -!- FUNCTION: Interacts with key regulators (CBP, p300 and PCAF) of
CC       transcription and represses transcription. Acts as a histone-
CC       binding protein that regulates transcription.
CC   -!- SUBUNIT: Interacts with DNA repair proteins RAD23A and RAD23B.
CC   -!- SUBCELLULAR LOCATION: Predominantly nuclear, but not exclusively;
CC       inner nuclear matrix.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC       Name=1;
CC         IsoId=P54252-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=P54252-2; Sequence=VSP_002784;
CC       Name=3;
CC         IsoId=P54252-3; Sequence=VSP_002783, VSP_002784;
CC   -!- TISSUE SPECIFICITY: Ubiquitous.
CC   -!- POLYMORPHISM: The poly-Gln region of the Machado-Joseph protein is
CC       highly polymorphic (14 to 41 repeats) in the normal population and
CC       is expanded to about 55-82 repeats in MJD1 patients. Longer
CC       expansions result in earlier onset and more severe clinical
CC       manifestations of the disease.
CC   -!- POLYMORPHISM: The MJD1a allele carries a single nucleotide
CC       substition in codon 349 generating a stop codon instead of a Tyr.
CC       In the Japanese population, the MJD1a allele seems to be
CC       significantly associated with Gln expansion.
CC   -!- DISEASE: Defects in MJD are the cause of Machado-Joseph disease
CC       (MJD), a neurodegenerative disorder characterized by cerebellar
CC       ataxia, pyramidal and extrapyramidal signs, peripheral nerve
CC       palsy, external ophtalmoplegia, facial and lingual fasciculation
CC       and bulging. This disease is autosomal and dominant, with a late
CC       onset of symptoms, generally after the fourth decade.
CC   -!- SIMILARITY: Contains 1 Josephin domain.
CC   -!- SIMILARITY: Contains 3 ubiquitin-interacting motif (UIM) repeats.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; S75313; AAB33571.1; -.
DR   EMBL; U64820; AAB63352.1; -.
DR   EMBL; U64821; AAB63353.1; -.
DR   EMBL; U64822; AAB63354.1; -.
DR   EMBL; AB050194; BAB18798.1; -.
DR   EMBL; AB038653; BAB55645.1; -.
DR   EMBL; AB038653; BAB55646.1; -.
DR   EMBL; BC033711; AAH33711.1; -.
DR   Genew; HGNC:7106; MJD.
DR   MIM; 607047; -.
DR   MIM; 109150; -.
DR   GO; GO:0005737; C:cytoplasm; TAS.
```

```
DR   GO; GO:0005654; C:nucleoplasm; TAS.
DR   GO; GO:0007399; P:neurogenesis; TAS.
DR   GO; GO:0006289; P:nucleotide-excision repair; TAS.
DR   GO; GO:0007268; P:synaptic transmission; TAS.
DR   InterPro; IPR006155; Josephin.
DR   InterPro; IPR003903; UIM.
DR   Pfam; PF02099; Josephin; 1.
DR   Pfam; PF02809; UIM; 2.
DR   PRINTS; PR01233; JOSEPHIN.
DR   SMART; SM00726; UIM; 2.
DR   PROSITE; PS50330; UIM; 2.
KW   Transcription regulation; Nuclear protein; Repeat;
KW   Alternative splicing; Polymorphism; Triplet repeat expansion.
FT   DOMAIN        1    198       JOSEPHIN.
FT   DOMAIN      224    243       UIM 1.
FT   DOMAIN      244    263       UIM 2.
FT   REPEAT      343    360       UIM 3.
FT   DOMAIN      292    317       POLY-GLN.
FT   VARSPLIC     10     64       Missing (in isoform 3).
FT                                /FTId=VSP_002783.
FT   VARSPLIC    344    376       KACSPFIMFATFTLYLTYELHVIFALHYSSFPL -> DAMS
FT                                EEDMLQAAVTMSLETVRNDLKTEGKK (in isoform 2
FT                                and isoform 3).
FT                                /FTId=VSP_002784.
FT   VARIANT     212    212       M -> V (in dbSNP:1048755).
FT                                /FTId=VAR_013688.
FT   VARIANT     306    318       QQQQQQQQQQQQR -> G.
FT                                /FTId=VAR_013689.
FT   VARIANT     361    376       Missing (in MJD1A).
FT                                /FTId=VAR_013690.
FT   CONFLICT    252    252       A -> T (IN REF. 2).
SQ   SEQUENCE   376 AA;  43449 MW;  C282BED37499480E CRC64;

  Query Match             47.7%;  Score 137;  DB 1;  Length 376;
  Best Local Similarity   68.2%;  Pred. No. 5.8e-06;
  Matches   30;  Conservative    4;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          5 GSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQ 48
              |:  | |:| |  |:    |||:|||||||||||||||||||||
Db        273 GTNLTSEELRKRREAYFEKQQQKQQQQQQQQQQQQQQQQQQQQ 316


RESULT 9
NCO6_MOUSE
ID   NCO6_MOUSE     STANDARD;      PRT;  2067 AA.
AC   Q9JL19; Q9JLT9;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Nuclear receptor coactivator 6 (Amplified in breast cancer-3 protein)
DE   (Cancer-amplified transcriptional coactivator ASC-2) (Activating
DE   signal cointegrator-2) (ASC-2) (Peroxisome proliferator-activated
DE   receptor-interacting protein) (PPAR-interacting protein) (Nuclear
DE   receptor-activating protein, 250 kDa) (Nuclear receptor coactivator
DE   RAP250) (NRC) (Thyroid hormone receptor binding protein).
GN   NCOA6 OR AIB3 OR RAP250 OR PRIP OR TRBP.
OS   Mus musculus (Mouse).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND INTERACTION WITH
RP   PPARA; PPARG; RARA; RXRA; ESR1; ESR2 AND THRB.
RC   TISSUE=Liver;
RX   MEDLINE=20250907; PubMed=10788465;
RA   Zhu Y.-J., Kan L., Qi C., Kanwar Y.S., Yeldandi A.V., Rao M.S.,
RA   Reddy J.K.;
RT   "Isolation and characterization of peroxisome proliferator-activated
RT   receptor (PPAR) interacting protein (PRIP) as a coactivator for
RT   PPAR.";
RL   J. Biol. Chem. 275:13510-13516(2000).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Breast;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 786-1142 FROM N.A. (ISOFORM 1), INTERACTION WITH PPARA;
RP   PPARG; ESR1; ESR2; THRA AND THRB, AND MUTAGENESIS OF LEU-891 AND
RP   LEU-894.
RC   TISSUE=Embryo;
RX   MEDLINE=20148724; PubMed=10681503;
RA   Caira F., Antonson P., Pelto-Huikko M., Treuter E., Gustafsson J.-A.;
RT   "Cloning and characterization of RAP250, a nuclear receptor
RT   coactivator.";
RL   J. Biol. Chem. 275:5308-5317(2000).
RN   [4]
RP   INTERACTION WITH RNPC2.
RX   MEDLINE=21638469; PubMed=11704680;
RA   Jung D.-J., Na S.-Y., Na D.S., Lee J.W.;
RT   "Molecular cloning and characterization of CAPER, a novel coactivator
RT   of activating protein-1 and estrogen receptors.";
RL   J. Biol. Chem. 277:1229-1234(2002).
CC   -!- FUNCTION: Nuclear receptor coactivator that directly binds nuclear
CC       receptors and stimulates the transcriptional activities in a
```

```
CC       hormone-dependent fashion. Coactivates expression in an agonist-
CC       and AF2-dependent manner. Involved in the coactivation of
CC       different nuclear receptors, such as for steroids (GR and ERs),
CC       retinoids (RARs and RXRs), thyroid hormone (TRs), vitamin D3 (VDR)
CC       and prostanoids (PPARs). Probably functions as a general
CC       coactivator, rather than just a nuclear receptor coactivator. May
CC       also be involved in the coactivation of the NF-kappa-B pathway.
CC       May coactivate expression via a remodeling of chromatin and its
CC       interaction with histone acetyltransferase proteins. Involved in
CC       placental, cardiac, hepatic and embryonic development.
CC   -!- SUBUNIT: Monomer and homodimer. Interacts in vitro with the basal
CC       transcription factors GTF2A and TBP, suggesting an autonomous
CC       transactivation function. Interacts with NCOA1, CRSP3, RBM14, the
CC       histone acetyltransferase proteins EP300 and CREBBP, and with
CC       methyltransferase proteins NCOA6IP and HRMT1L1 (By similarity).
CC       Interacts with RNPC2. Belongs to the ASC-2/NCOA6 complex (ASCOM),
CC       which contains ASC-2/NCOA6, the retinoblastoma-binding protein
CC       RBQ-3/ RBBP5, alpha- and beta-tubulins, the trithorax group
CC       proteins MLL2 and MLL3, and ASH2/ASCL2 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q9JL19-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q9JL19-2; Sequence=VSP_003410;
CC         Note=Acts as a dominant negative repressor;
CC   -!- TISSUE SPECIFICITY: Widely expressed. High expression in testis
CC       and weak expression in small intestine.
CC   -!- DEVELOPMENTAL STAGE: Expressed at E9 in placenta and at weaker
CC       level in uterus. High expression in neural tube and in CNS
CC       throughout development. High expression in sensory ganglia and
CC       retina from E11. In the alimentary tract and olfactory epithelium
CC       expression was seen from E13. Strong expression present in liver
CC       and kidney, from E11 and E13 respectively, and then expression
CC       decreased at later stages of development. Moderate expression in
CC       lung from E13, while it decreases during postnatal life. Strong
CC       expression in thymus from E15 onwards, and in spleen from E17 and
CC       during early postnatal life, then, the expression decreases.
CC   -!- DOMAIN: Contains two Leu-Xaa-Xaa-Leu-Leu (LXXLL) motifs. Only
CC       motif 1 is essential for the association with nuclear receptors.
CC   -!- PTM: Phosphorylated (By similarity).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF216186; AAF35860.1; -.
DR   EMBL; BC031113; AAH31113.1; -.
DR   EMBL; AF135169; AAF35973.1; -.
DR   MGD; MGI:1929915; Ncoa6.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0005667; C:transcription factor complex; IDA.
```

```
DR   GO; GO:0003682; F:chromatin binding; IDA.
DR   GO; GO:0030331; F:estrogen receptor binding; ISS.
DR   GO; GO:0046965; F:retinoid X receptor binding; ISS.
DR   GO; GO:0046966; F:thyroid hormone receptor binding; ISS.
DR   GO; GO:0003713; F:transcription co-activator activity; ISS.
DR   GO; GO:0016563; F:transcriptional activator activity; IDA.
DR   GO; GO:0007420; P:brain development; IMP.
DR   GO; GO:0001701; P:embryonic development (sensu Mammalia); IMP.
DR   GO; GO:0007507; P:heart development; IMP.
DR   GO; GO:0030099; P:myeloid blood cell differentiation; ISS.
DR   GO; GO:0006367; P:transcription initiation from Pol II promoter; IDA.
KW   Transcription regulation; Activator; Nuclear protein; Repeat;
KW   Alternative splicing.
FT   DOMAIN        1   1060       CREBBP-BINDING REGION (BY SIMILARITY).
FT   DOMAIN        1    932       TBP/GTF2A-BINDING REGION (BY SIMILARITY).
FT   DOMAIN        1   1314       NCOA1-BINDING REGION (BY SIMILARITY).
FT   DOMAIN      777    931       NCOA6IP-BINDING REGION (BY SIMILARITY).
FT   DOMAIN     1644   2067       EP300/CRSP3-BINDING REGION
FT                                (BY SIMILARITY).
FT   DOMAIN      227   1044       GLN-RICH.
FT   DOMAIN      376    381       POLY-PRO.
FT   DOMAIN      917    922       POLY-LYS.
FT   DOMAIN     1543   1592       SER-RICH.
FT   SITE        891    895       LXXLL MOTIF 1.
FT   SITE       1495   1499       LXXLL MOTIF 2.
FT   VARSPLIC    458   2067       Missing (in isoform 2).
FT                                /FTId=VSP_003410.
FT   MUTAGEN     891    894       LVNL->AVNA: ABOLISHES INTERACTION WITH
FT                                NUCLEAR RECEPTORS.
FT   CONFLICT     39     39       G -> S (IN REF. 2).
FT   CONFLICT    109    109       W -> R (IN REF. 2).
FT   CONFLICT    194    194       M -> I (IN REF. 2).
FT   CONFLICT    290    290       Q -> QQ (IN REF. 2).
FT   CONFLICT   1014   1014       P -> L (IN REF. 3).
FT   CONFLICT   1141   1142       SE -> RS (IN REF. 3).
SQ   SEQUENCE   2067 AA;  219663 MW;  C855F8777167AD48 CRC64;

  Query Match             47.6%;  Score 136.5;  DB 1;  Length 2067;
  Best Local Similarity   54.2%;  Pred. No. 3e-05;
  Matches   32;  Conservative    3;  Mismatches   13;  Indels   11;  Gaps    1;

Qy          3 PRGSMATLEKLMKA-----------FESLKSFQQQQQQQQQQQQQQQQQQQQQQQQLQ 50
              | ||:      |:            |  |:  |||||||||||||||||||||||| |
Db        232 PSGSLPPAHHSMQPVPVNRQMNPANFPQLQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 290


RESULT 10
THAB_HUMAN
ID   THAB_HUMAN     STANDARD;      PRT;   313 AA.
AC   Q96EK4; O94795;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   THAP domain protein 11 (HRIHFB2206).
GN   THAP11.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Uterus;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE OF 226-313 FROM N.A., AND SUBCELLULAR LOCATION.
RC   TISSUE=Brain;
RX   MEDLINE=99068504; PubMed=9853615;
RA   Ueki N., Oda T., Kondo M., Yano K., Noguchi T., Muramatsu M.A.;
RT   "Selection system for genes encoding nuclear-targeted proteins.";
RL   Nat. Biotechnol. 16:1338-1342(1998).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: Contains 1 THAP domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; BC012182; AAH12182.1; -.
DR   EMBL; AB015338; BAA34796.1; -.
DR   Genew; HGNC:23194; THAP11.
DR   InterPro; IPR006612; DUF_DM3.
DR   Pfam; PF05485; THAP; 1.
DR   SMART; SM00692; DM3; 1.
KW   Zinc-finger; DNA-binding; Nuclear protein.
FT   DOMAIN        1     85       THAP.
FT   ZN_FING       6     64       THAP-TYPE.
FT   DOMAIN       95    100       POLY-ALA.
FT   DOMAIN      104    131       POLY-GLN.
FT   DOMAIN      201    207       POLY-ALA.
SQ   SEQUENCE   313 AA;  34327 MW;  47D8B02FF89E5BEB CRC64;
```

```
  Query Match             47.4%;  Score 136;  DB 1;  Length 313;
  Best Local Similarity   55.6%;  Pred. No. 6e-06;
  Matches   30;  Conservative    5;  Mismatches   11;  Indels    8;  Gaps    1;

Qy          3 PRGSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQLQPGSTRA 56
              | |: |   :  :        ||||||||||||||||||||||||  | :: |
Db         94 PAGAAAARRRQQQ--------QQQQQQQQQQQQQQQQQQQQQQQQSSPSASTA 139


RESULT 11
FXP2_MOUSE
ID   FXP2_MOUSE     STANDARD;      PRT;   714 AA.
AC   P58463;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Forkhead box protein P2.
GN   FOXP2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Lung;
RX   MEDLINE=21347947; PubMed=11358962;
RA   Shu W., Yang H., Zhang L., Lu M.M., Morrisey E.E.;
RT   "Characterization of a new subfamily of winged-helix/forkhead (Fox)
RT   genes that are expressed in the lung and act as transcriptional
RT   repressors.";
RL   J. Biol. Chem. 276:27488-27497(2001).
CC   -!- FUNCTION: Transcriptional repressor that play an important role in
CC       the specification and differentiation of lung epithelium. May play
CC       important roles in developing neural, gastrointestinal and
CC       cardiovascular tissues.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: Highest expression in lung. Lower expression
CC       in spleen, skeletal muscle, brain, kidney and small intestine.
CC   -!- DEVELOPMENTAL STAGE: Expressed in developing lung (only distal
CC       epithelium), neural, intestinal and cardiovascular tissues.
CC   -!- SIMILARITY: Contains 1 fork-head domain.
CC   -!- SIMILARITY: Contains 1 C2H2-type zinc finger.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF339106; AAK69651.1; -.
DR   MGD; MGI:2148705; Foxp2.
DR   GO; GO:0016564; F:transcriptional repressor activity; IDA.
DR   GO; GO:0016481; P:negative regulation of transcription; IDA.
DR   InterPro; IPR001766; TF_Fork_head.
```

```
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00250; Fork_head; 1.
DR   PRINTS; PR00053; FORKHEAD.
DR   ProDom; PD000425; TF_Fork_head; 1.
DR   SMART; SM00339; FH; 1.
DR   SMART; SM00355; ZnF_C2H2; 1.
DR   PROSITE; PS00657; FORK_HEAD_1; FALSE_NEG.
DR   PROSITE; PS00658; FORK_HEAD_2; FALSE_NEG.
DR   PROSITE; PS50039; FORK_HEAD_3; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; FALSE_NEG.
KW   Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW   Nuclear protein.
FT   ZN_FING     345    370       C2H2-TYPE.
FT   DNA_BIND    503    593       FORK-HEAD.
FT   DOMAIN       53     56       POLY-GLN.
FT   DOMAIN      123    126       POLY-GLN.
FT   DOMAIN      131    136       POLY-GLN.
FT   DOMAIN      152    191       POLY-GLN.
FT   DOMAIN      200    208       POLY-GLN.
FT   DOMAIN      222    230       POLY-GLN.
SQ   SEQUENCE   714 AA;  79820 MW;  BCDFB80E28398609 CRC64;

  Query Match             47.4%;  Score 136;  DB 1;  Length 714;
  Best Local Similarity   93.1%;  Pred. No. 1.3e-05;
  Matches   27;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         24 QQQQQQQQQQQQQQQQQQQQQQQQQLQPG 52
              ||||||||||||||||||||||||| ||
Db        166 QQQQQQQQQQQQQQQQQQQQQQQQQHPG 194


RESULT 12
FXP2_HUMAN
ID   FXP2_HUMAN     STANDARD;      PRT;   715 AA.
AC   O15409; Q8N0W2;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Forkhead box protein P2 (CAG repeat protein 44) (Trinucleotide repeat-
DE   containing gene 10 protein).
GN   FOXP2 OR CAGH44 OR TNRC10.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND VARIANT SPCH1 HIS-553.
RX   MEDLINE=21470412; PubMed=11586359;
RA   Lai C.S.L., Fisher S.E., Hurst J.A., Vargha-Khadem F., Monaco A.P.;
RT   "A forkhead-domain gene is mutated in a severe speech and language
RT   disorder.";
RL   Nature 413:519-523(2001).
RN   [2]
RP   SEQUENCE OF 1-304 FROM N.A.
RC   TISSUE=Brain cortex;
RX   MEDLINE=97369492; PubMed=9225980;
```

```
RA   Margolis R.L., Abraham M.R., Gatchell S.B., Li S.-H., Kidwai A.S.,
RA   Breschel T.S., Stine O.C., Callahan C., Mcinnis M.G., Ross C.A.;
RT   "cDNAs with long CAG trinucleotide repeats from human brain.";
RL   Hum. Genet. 100:114-122(1997).
RN   [3]
RP   SEQUENCE OF 1-86 FROM N.A.
RA   Minx P., Hinds K., Sutterer C., Becker M., Ozersky P.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 113-329 FROM N.A.
RX   MEDLINE=22179809; PubMed=12192408;
RA   Enard W., Przeworski M., Fisher S.E., Lai C.S.L., Wiebe V., Kitano T.,
RA   Monaco A.P., Paabo S.;
RT   "Molecular evolution of FOXP2, a gene involved in speech and
RT   language.";
RL   Nature 418:869-872(2002).
CC   -!- FUNCTION: Transcriptional repressor that plays an important role
CC       in the specification and differentiation of lung epithelium. May
CC       play important roles in developing neural, gastrointestinal and
CC       cardiovascular tissues. Involved in neural mechanisms mediating
CC       the development of speech and language.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC       Name=1; Synonyms=I;
CC         IsoId=O15409-1; Sequence=Displayed;
CC       Name=2; Synonyms=II;
CC         IsoId=O15409-3; Sequence=Not described;
CC       Name=3; Synonyms=III, IV;
CC         IsoId=O15409-2; Sequence=VSP_001558;
CC   -!- TISSUE SPECIFICITY: Expressed at high levels in embryonic and
CC       adult lung.
CC   -!- DISEASE: Defects in FOXP2 are the cause of speech-language
CC       disorder 1 (SPCH1) [MIM:602081]; also known as autosomal dominant
CC       speech and language disorder with orofacial dyspraxia. Affected
CC       individuals have a severe impairment in the selection and
CC       sequencing of fine orofacial movements, which are necessary for
CC       articulation. They also show deficits in several facets of
CC       language processing (such as the ability to break up words into
CC       their constituent phoneme) and grammatical skills.
CC   -!- DISEASE: Disruption of FOXP2 by a chromosomal translocation
CC       t(5;7)(q22;q31.2) is the cause of severe speech and language
CC       impairment.
CC   -!- SIMILARITY: Contains 1 fork-head domain.
CC   -!- SIMILARITY: Contains 1 C2H2-type zinc finger.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF337817; AAL10762.1; -.
DR   EMBL; U80741; AAB91439.1; -.
DR   EMBL; AC003992; -; NOT_ANNOTATED_CDS.
```

```
DR   EMBL; AF515031; AAN03389.1; -.
DR   EMBL; AF515032; AAN03390.1; -.
DR   EMBL; AF515033; AAN03391.1; -.
DR   EMBL; AF515034; AAN03392.1; -.
DR   EMBL; AF515035; AAN03393.1; -.
DR   EMBL; AF515036; AAN03394.1; -.
DR   EMBL; AF515037; AAN03395.1; -.
DR   EMBL; AF515038; AAN03396.1; -.
DR   EMBL; AF515039; AAN03397.1; -.
DR   EMBL; AF515040; AAN03398.1; -.
DR   EMBL; AF515041; AAN03399.1; -.
DR   EMBL; AF515042; AAN03400.1; -.
DR   EMBL; AF515043; AAN03401.1; -.
DR   EMBL; AF515044; AAN03402.1; -.
DR   EMBL; AF515045; AAN03403.1; -.
DR   EMBL; AF515046; AAN03404.1; -.
DR   EMBL; AF515047; AAN03405.1; -.
DR   EMBL; AF515048; AAN03406.1; -.
DR   EMBL; AF515049; AAN03407.1; -.
DR   EMBL; AF515050; AAN03408.1; -.
DR   Genew; HGNC:13875; FOXP2.
DR   MIM; 605317; -.
DR   MIM; 602081; -.
DR   InterPro; IPR001766; TF_Fork_head.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00250; Fork_head; 1.
DR   PRINTS; PR00053; FORKHEAD.
DR   ProDom; PD000425; TF_Fork_head; 1.
DR   SMART; SM00339; FH; 1.
DR   SMART; SM00355; ZnF_C2H2; 1.
DR   PROSITE; PS00657; FORK_HEAD_1; FALSE_NEG.
DR   PROSITE; PS00658; FORK_HEAD_2; FALSE_NEG.
DR   PROSITE; PS50039; FORK_HEAD_3; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; FALSE_NEG.
KW   Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW   Nuclear protein; Chromosomal translocation; Disease mutation;
KW   Alternative splicing.
FT   ZN_FING      346    371       C2H2-TYPE.
FT   DNA_BIND     504    594       FORK-HEAD.
FT   DOMAIN        53     56       POLY-GLN.
FT   DOMAIN       123    126       POLY-GLN.
FT   DOMAIN       131    136       POLY-GLN.
FT   DOMAIN       152    191       POLY-GLN.
FT   DOMAIN       200    209       POLY-GLN.
FT   DOMAIN       223    231       POLY-GLN.
FT   VARSPLIC       1     92       Missing (in isoform 3).
FT                                  /FTId=VSP_001558.
FT   VARIANT      553    553       R -> H (in SPCH1).
FT                                  /FTId=VAR_012278.
FT   CONFLICT     134    134       Q -> H (IN REF. 2).
FT   CONFLICT     290    304       DLTTNNSSSTTSSNT -> EEFPVQGPAAVCAGL (IN
FT                                  REF. 2).
SQ   SEQUENCE   715 AA;  79919 MW;  4F9FBDB6D90516E0 CRC64;

  Query Match             47.4%;  Score 136;  DB 1;  Length 715;
  Best Local Similarity   93.1%;  Pred. No. 1.3e-05;
```

```
  Matches   27;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         24 QQQQQQQQQQQQQQQQQQQQQQQQQLQPG 52
              ||||||||||||||||||||||||||  ||
Db        166 QQQQQQQQQQQQQQQQQQQQQQQQQHPG 194


RESULT 13
FXP2_PANTR
ID   FXP2_PANTR     STANDARD;      PRT;   716 AA.
AC   Q8MJA0; Q8MHX3;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Forkhead box protein P2.
GN   FOXP2.
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX   NCBI_TaxID=9598;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22179809; PubMed=12192408;
RA   Enard W., Przeworski M., Fisher S.E., Lai C.S.L., Wiebe V., Kitano T.,
RA   Monaco A.P., Paabo S.;
RT   "Molecular evolution of FOXP2, a gene involved in speech and
RT   language.";
RL   Nature 418:869-872(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22412141; PubMed=12524352;
RA   Zhang J., Webb D.M., Podlaha O.;
RT   "Accelerated protein evolution and origins of human-specific features:
RT   Foxp2 as an example.";
RL   Genetics 162:1825-1835(2002).
CC   -!- FUNCTION: Transcriptional repressor that plays an important role
CC       in the specification and differentiation of lung epithelium. May
CC       play important roles in developing neural, gastrointestinal and
CC       cardiovascular tissues (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: Contains 1 fork-head domain.
CC   -!- SIMILARITY: Contains 1 C2H2-type zinc finger.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF512947; AAN03385.1; -.
DR   EMBL; AF515051; AAN03409.1; -.
DR   EMBL; AF515052; AAN03410.1; -.
DR   EMBL; AY143178; AAN60056.1; -.
DR   InterPro; IPR001766; TF_Fork_head.
DR   InterPro; IPR009058; Wing_hlx_DNA_bnd.
```

```
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00250; Fork_head; 1.
DR   PRINTS; PR00053; FORKHEAD.
DR   ProDom; PD000425; TF_Fork_head; 1.
DR   SMART; SM00339; FH; 1.
DR   PROSITE; PS00657; FORK_HEAD_1; FALSE_NEG.
DR   PROSITE; PS00658; FORK_HEAD_2; FALSE_NEG.
DR   PROSITE; PS50039; FORK_HEAD_3; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; FALSE_NEG.
KW   Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW   Nuclear protein.
FT   ZN_FING     347    372       C2H2-TYPE.
FT   DNA_BIND    505    595       FORK-HEAD.
FT   DOMAIN       53     56       POLY-GLN.
FT   DOMAIN      123    126       POLY-GLN.
FT   DOMAIN      131    136       POLY-GLN.
FT   DOMAIN      152    191       POLY-GLN.
FT   DOMAIN      201    210       POLY-GLN.
FT   DOMAIN      224    232       POLY-GLN.
SQ   SEQUENCE   716 AA;  80061 MW;  3169A2786B42F79F CRC64;

  Query Match             47.4%;  Score 136;  DB 1;  Length 716;
  Best Local Similarity   93.1%;  Pred. No. 1.3e-05;
  Matches   27;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         24 QQQQQQQQQQQQQQQQQQQQQQQQQLQPG 52
              ||||||||||||||||||||||||| ||
Db        167 QQQQQQQQQQQQQQQQQQQQQQQQQHPG 195


RESULT 14
MN1_HUMAN
ID   MN1_HUMAN      STANDARD;      PRT;  1319 AA.
AC   Q10571;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Probable tumor suppressor protein MN1.
GN   MN1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=95249266; PubMed=7731706;
RA   Deprez R.H.L., Riegman P.H.J., Groen N.A., Warringa U.L.,
RA   van Biezen N.A., Molijn A.C., Bootsma D., de Jong P.J.,
RA   Menon A.G., Kley N.A., Seizenger B.R., Zwarthoff E.C.;
RT   "Cloning and characterization of MN1, a gene from chromosome 22q11,
RT   which is disrupted by a balanced translocation in a meningioma.";
RL   Oncogene 10:1521-1528(1995).
RN   [2]
RP   SEQUENCE OF 1304-1319 FROM N.A.
RC   TISSUE=Brain;
```

```
RX   MEDLINE=97145634; PubMed=9026990;
RA   Dmitrenko V.V., Garifulin O.M., Shostak E.A., Smikodub A.I.,
RA   Kavsan V.M.;
RT   "The characteristics of different types of mRNA expressed in the human
RT   brain.";
RL   Cyt. Genet. (Russ.) 30:41-47(1996).
CC   -!- FUNCTION: May play a role in tumor suppression.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q10571-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q10571-2; Sequence=Not described;
CC         Note=No experimental confirmation available;
CC   -!- TISSUE SPECIFICITY: Ubiquitously expressed. Highest levels in
CC       skeletal muscle.
CC   -!- DISEASE: Involved in a form of acute myeloid leukemia (AML) by a
CC       chromosomal translocation t(12;22)(p13;q11) that involves MN1 and
CC       TEL.
CC   -!- DISEASE: Defects in MN1 may be a cause of meningiomas, slowly
CC       growing benign tumors derived from the arachnoidal cap cells of
CC       the leptomeninges, the soft coverings of the brain and spinal
CC       cord. Meningiomas are believed to be the most common primary
CC       tumors of the central nervous system in man.
CC   -!- CAUTION: It is uncertain whether Met-1 or Met-30 is the initiator.
CC   -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC       WWW="http://www.infobiogen.fr/services/chromcancer/Genes/MN1.html".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X82209; CAA57693.1; ALT_INIT.
DR   EMBL; Z70218; CAA94179.1; -.
DR   Genew; HGNC:7180; MN1.
DR   MIM; 156100; -.
DR   MIM; 607174; -.
KW   Anti-oncogene; Chromosomal translocation; Alternative splicing.
FT   DOMAIN      295    309       POLY-GLN.
FT   DOMAIN      523    550       POLY-GLN.
SQ   SEQUENCE   1319 AA;  135943 MW;  21197C9BBA272BE2 CRC64;

  Query Match             47.4%;  Score 136;  DB 1;  Length 1319;
  Best Local Similarity   84.8%;  Pred. No. 2.2e-05;
  Matches   28;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         18 ESLKSFQQQQQQQQQQQQQQQQQQQQQQQQLQ 50
              :||:  |||||||||||||||||||||||| |
Db        520 QSLQQQQQQQQQQQQQQQQQQQQQQQQQQQRQ 552


RESULT 15
HCN1_MOUSE
```

```
ID   HCN1_MOUSE     STANDARD;      PRT;   910 AA.
AC   O88704; O54899; Q9D613;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Potassium/sodium hyperpolarization-activated cyclic nucleotide-gated
DE   channel 1 (Brain cyclic nucleotide gated channel 1) (BCNG-1)
DE   (Hyperpolarization-activated cation channel 2) (HAC-2).
GN   HCN1 OR BCNG1 OR HAC2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND N-GLYCOSYLATION.
RC   STRAIN=C57BL/6J; TISSUE=Brain;
RX   MEDLINE=98070835; PubMed=9405696;
RA   Santoro B., Grant S.G.N., Bartsch D., Kandel E.R.;
RT   "Interactive cloning with the SH3 domain of N-src identifies a new
RT   brain specific ion channel protein, with homology to eag and cyclic
RT   nucleotide-gated channels.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:14815-14820(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Brain;
RX   MEDLINE=98295993; PubMed=9634236;
RA   Ludwig A., Zong X., Jeglitsch M., Hofmann F., Biel M.;
RT   "A family of hyperpolarization-activated cation channels.";
RL   Nature 393:587-591(1998).
RN   [3]
RP   SEQUENCE OF 377-910 FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Head;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [4]
RP   FUNCTION, AND REGULATION BY CAMP.
RX   MEDLINE=98292171; PubMed=9630217;
RA   Santoro B., Liu D.T., Yao H., Bartsch D., Kandel E.R.,
```

RA Siegelbaum S.A., Tibbs G.R.;
RT "Identification of a gene encoding a hyperpolarization-activated
RT pacemaker channel of brain.";
RL Cell 93:717-729(1998).
RN [5]
RP INTERACTION WITH KCNE2.
RX MEDLINE=21313430; PubMed=11420311;
RA Yu H., Wu J., Potapova I., Wymore R.T., Holmes B., Zuckerman J.,
RA Pan Z., Wang H., Shi W., Robinson R.B., El-Maghrabi M.R., Benjamin W.,
RA Dixon J.E., McKinnon D., Cohen I.S., Wymore R.;
RT "MinK-related peptide 1: A beta subunit for the HCN ion channel
RT subunit family enhances expression and speeds activation.";
RL Circ. Res. 88:E84-E87(2001).
RN [6]
RP REGULATION BY CAMP.
RX MEDLINE=21351681; PubMed=11459060;
RA Wainger B.J., DeGennaro M., Santoro B., Siegelbaum S.A., Tibbs G.R.;
RT "Molecular mechanism of cAMP modulation of HCN pacemaker channels.";
RL Nature 411:805-810(2001).
RN [7]
RP FUNCTION, AND TISSUE SPECIFICITY.
RX MEDLINE=21530492; PubMed=11675786;
RA Stevens D.R., Seifert R., Bufe B., Mueller F., Kremmer E., Gauss R.,
RA Meyerhof W., Kaupp U.B., Lindemann B.;
RT "Hyperpolarization-activated channels HCN1 and HCN4 mediate responses
RT to sour stimuli.";
RL Nature 413:631-635(2001).
RN [8]
RP INTERACTION WITH HCN2, AND MUTAGENESIS OF GLY-349; TYR-350 AND
RP GLY-351.
RX MEDLINE=22083667; PubMed=12089064;
RA Xue T., Marban E., Li R.A.;
RT "Dominant-negative suppression of HCN1- and HCN2-encoded pacemaker
RT currents by an engineered HCN1 construct: insights into
RT structure-function relationships and multimerization.";
RL Circ. Res. 90:1267-1273(2002).
RN [9]
RP OLIGOMERIZATION VIA N-TERMINAL DOMAIN.
RX MEDLINE=22162449; PubMed=12034718;
RA Proenza C., Tran N., Angoli D., Zahynacz K., Balcar P., Accili E.A.;
RT "Different roles for the cyclic nucleotide binding domain and amino
RT terminus in assembly and expression of hyperpolarization-activated,
RT cyclic nucleotide-gated channels.";
RL J. Biol. Chem. 277:29634-29642(2002).
RN [10]
RP MUTAGENESIS OF CYS-303 AND CYS-318.
RX MEDLINE=22336443; PubMed=12351622;
RA Xue T., Li R.A.;
RT "An external determinant in the S5-P linker of the pacemaker (HCN)
RT channel identified by sulfhydryl modification.";
RL J. Biol. Chem. 277:46233-46242(2002).
CC -!- FUNCTION: Hyperpolarization-activated ion channel exhibiting weak
CC selectivity for potassium over sodium ions. Contributes to the
CC native pacemaker currents in heart (If) and in neurons (Ih).
CC Activated by cAMP, and at 10-100 times higher concentrations, also
CC by cGMP. May mediate responses to sour stimuli.
CC -!- SUBUNIT: The potassium channel is probably composed of a homo- or

```
CC       heterotetrameric complex of pore-forming subunits. Heteromultimer
CC       with HCN2. Interacts with KCNE2. Interacts with the SH3 domain of
CC       CSK.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Predominantly expressed in brain. Highly
CC       expressed in apical dendrites of pyramidal neurons in the cortex,
CC       in the layer corresponding to the stratum lacunosum-moleculare in
CC       the hippocampus and in axons of basket cells in the cerebellum.
CC       Expressed in a subset of elongated cells in taste buds.
CC   -!- DOMAIN: The segment S4 is probably the voltage-sensor and is
CC       characterized by a series of positively charged amino acids at
CC       every third position.
CC   -!- PTM: N-glycosylated.
CC   -!- MISCELLANEOUS: Inhibited by extracellular cesium ions.
CC   -!- SIMILARITY: Belongs to the potassium channel family. HCN
CC       subfamily.
CC   -!- SIMILARITY: Contains 1 cyclic nucleotide-binding domain.
CC   -!- CAUTION: Ref.3 sequence differs from that shown due to a
CC       frameshift in position 381.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF028737; AAC53518.1; -.
DR   EMBL; AJ225123; CAA12407.1; -.
DR   EMBL; AK014722; BAB29519.1; ALT_FRAME.
DR   MGD; MGI:1096392; Hcn1.
DR   InterPro; IPR000595; cNMP_binding.
DR   InterPro; IPR005821; Ion_trans.
DR   InterPro; IPR001622; K+channel_pore.
DR   InterPro; IPR005820; M+channel_nlg.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; FALSE_NEG.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
KW   Transport; Ion transport; Ionic channel; Voltage-gated channel;
KW   Potassium channel; Potassium; Potassium transport; Sodium transport;
KW   cAMP; cAMP-binding; Transmembrane; Glycoprotein; Sodium channel.
FT   DOMAIN        1    135       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    136    156       SEGMENT S1 (POTENTIAL).
FT   TRANSMEM    163    183       SEGMENT S2 (POTENTIAL).
FT   DOMAIN      184    208       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    209    229       SEGMENT S3 (POTENTIAL).
FT   TRANSMEM    238    258       SEGMENT S4 (POTENTIAL).
FT   DOMAIN      259    289       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    290    310       SEGMENT S5 (POTENTIAL).
FT   TRANSMEM    334    355       SEGMENT H5 (PORE-FORMING) (POTENTIAL).
FT   TRANSMEM    361    381       SEGMENT S6 (POTENTIAL).
FT   DOMAIN      382    910       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       78    129       INVOLVED IN SUBUNIT ASSEMBLY (BY
```

```
FT                                     SIMILARITY).
FT   NP_BIND      464    581           CAMP.
FT   DOMAIN         1     81           GLY-RICH.
FT   DOMAIN       715    777           GLN-RICH.
FT   DOMAIN       878    884           POLY-PRO.
FT   CARBOHYD     327    327           N-LINKED (GLCNAC. . .) (PROBABLE).
FT   MUTAGEN      303    303           C->S: ABOLISHES CONDUCTIVITY.
FT   MUTAGEN      318    318           C->S: ABOLISHES SENSITIVITY TO SULFHYDRIL
FT                                     MODIFICATION.
FT   MUTAGEN      349    349           G->A: ABOLISHES CONDUCTIVITY; WHEN
FT                                     ASSOCIATED WITH A-350 AND A-351.
FT   MUTAGEN      350    350           Y->A: ABOLISHES CONDUCTIVITY; WHEN
FT                                     ASSOCIATED WITH A-349 AND A-351.
FT   MUTAGEN      351    351           G->A: ABOLISHES CONDUCTIVITY; WHEN
FT                                     ASSOCIATED WITH A-349 AND A-350.
FT   CONFLICT      42     42           G -> R (IN REF. 1).
FT   CONFLICT     394    394           R -> S (IN REF. 3).
SQ   SEQUENCE   910 AA;  102432 MW;  56FD5F328DD972E9 CRC64;

  Query Match             47.0%;  Score 135;  DB 1;  Length 910;
  Best Local Similarity   96.4%;  Pred. No. 2e-05;
  Matches   27;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         24 QQQQQQQQQQQQQQQQQQQQQQQQQLQP 51
              ||||||||||||||||||||||||| ||
Db        749 QQQQQQQQQQQQQQQQQQQQQQQQQQP 776


Search completed: March 12, 2004, 15:39:06
Job time : 7.94118 secs
```